Flexible visual information representation in human parietal cortex

Flexible visual information representation in human parietal cortex

A dissertation presented
By
Su Keun Jeong
To
The Department of Psychology

In partial fulfillment of the requirements
For the degree of
Doctor of Philosophy
In the subject of
Psychology

Harvard University
Cambridge, Massachusetts
September 2014

Dissertation advisor: Professor Yaoda Xu Su Keun Jeong

Flexible visual information representation in human parietal cortex

Abstract

In many everyday activities, we must visually process multiple objects embedded in complex real world scenes. Our visual system can flexibly extract behaviorally relevant visual information from such scenes, even though it has a severely limited processing capacity. This dissertation proposes that human superior intra-parietal sulcus (IPS) plays a central role in this flexible visual information processing. In Chapter 1, using functional magnetic resonance imaging (fMRI) with univariate analysis, I found that distractor processing in superior IPS was attenuated when target locations were known in advance. In Chapter 2, using multi-voxel pattern analysis (MVPA), I showed that superior IPS encoded object shapes, but only when such information was required by task. In Chapter 3, I showed that, given a set of perceptually distinct, but semantically grouped visual inputs, superior IPS could represent abstract object identity. The neural similarity of identities in superior IPS significantly correlated with perceived similarity between identities, confirming the representation in this region indeed reflected identity. Taken together, these results suggest that human superior IPS encodes a wide range of visual information, from simple features to abstract identities, in a task-dependent manner, enabling flexible goal-directed visual information processing in the human brain.

Table of Contents

Acknowledgements

I would like to express my sincere gratitude to my advisor Yaoda Xu who supported and guided me with her patience. I would also like to thank my dissertation committee members: Ken Nakayama, George Alvarez and Alfonso Caramazza.

Thank you to my advisors at Yonsei University, Min-Shik Kim, Sang Chul Chong, Do-Joon Yi and Chansup Chung, for the encouragement and guidance that they have given to me.

Many thanks go to the members of Harvard Vision Sciences Laboratory and psychology department cohort.

Special thanks go to ELP friends who provided me with emotional support, fun and (lots of) food.

Lastly, I would like to offer my appreciation to my family for their unconditional support throughout my whole life.

This work was supported by the National Science Foundation grant 0855112 and the National Institute of Health grant 1R01EY022355 to Yaoda Xu, and Harvard University GSAS Dissertation Completion Fellowship.

0

Introduction

0.1 Capacity limits of the visual system and flexible information processing

In everyday life, our visual system often has to process multiple objects embedded in complex visual scenes and extract what is relevant for the current goal. However, the processing capacity of the visual system is severely limited (Marois & Ivanoff, 2005). For example, visual short-term memory (VSTM) (Baddeley, 1986; Phillips, 1974) can only encode a small amount of information, roughly about four objects worth, at a time (Alvarez & Cavanagh, 2004; Cowan, 2001; Luck & Vogel, 1997; Zhang & Luck, 2008). To overcome

these limitations, the visual system extracts task-relevant information from complex visual scenes while ignoring task-irrelevant information. The selection of the relevant information and the filtering of the distractors can occur early on during information processing, thus preventing the processing of distractors all together (Broadbent, 1958; Moray, 1959), or, it can occur later, such that all incoming information is processed to some degree, but information related to distractors is subsequently discarded (J. A. Deutsch & Deutsch, 1963; Luck, Vogel, & Shapiro, 1996). Our visual system also flexibly modulates task-irrelevant information processing depending on processing load (Jeong & Xu, 2013; Lavie, 2005; Lavie & Tsal, 1994; Lavie, Hirst, de Fockert, & Viding, 2004; Yi, Woodman, Widders, Marois, & Chun, 2004). In this case, task-irrelevant information will be discarded at the early stage when either the perceptual load is high or when the cognitive control mechanism is available to reject distractors (Lavie, 2005; Lavie et al., 2004). Additionally, the selection of relevant information can be either object- or feature-based. Consider, for example, trying to find a cab in a busy street. You will ignore all the distracting billboards and buildings next to the street because they are not relevant to your current goal. Instead, you will focus only the vehicles on the street. If you know all the cabs have a certain color (e.g., yellow), you might even encode only the color of the vehicles without processing all the detailed features. Supporting this, previous research has shown that items can be selected for further processing when they are task-relevant (Jeong & Xu, 2013; O'Craven, Downing, & Kanwisher, 1999; Scholl, 2001) or contain a task-relevant feature(s) (Corbetta, Miezin, Dobmeyer, Shulman, & Petersen, 1990; Serences, Ester, Vogel, & Awh, 2009; Xu, 2010; Xu & Jeong, in press). Researchers have suggested a network of brain regions including frontal

and parietal cortices is involved in such flexible information processing (Cole et al., 2013; Dosenbach, Fair, Cohen, Schlaggar, & Petersen, 2008; Duncan, 2001; 2010; Miller & Cohen, 2001; Vincent, Kahn, Snyder, Raichle, & Buckner, 2008), but relatively little is known about the role of parietal cortex in this network. This dissertation investigated the neural mechanisms in human parietal cortex that support flexible visual information processing.

0.2 Visual information processing in the primate brain

Visual information processing in the primate brain is commonly thought to involve two anatomically and functionally distinct pathways (Ungerleider & Mishkin, 1982). In this two-pathway view, the occipitotemporal, or ventral, pathway is involved in object perception and recognition (Kravitz, Saleem, Baker, Ungerleider, & Mishkin, 2013; Ungerleider & Haxby, 1994; Ungerleider & Mishkin, 1982). The occipitoparietal, or dorsal, pathway is involved in spatial vision as well as motor actions directed to objects (Goodale & Milner, 1992; Kravitz, Saleem, Baker, & Mishkin, 2011; Ungerleider & Haxby, 1994; Ungerleider & Mishkin, 1982).

However, growing evidence suggests that other aspects of visual information, and not only location and action, can be represented in the dorsal pathway. Specifically, studies in both monkeys and humans have identified a sub-region in IPS that encodes visual object information in a task-dependent manner. In monkeys, lateral intra-parietal (LIP) neurons have been found to show selectivity for non-spatial information such as shape and color (Sereno & Maunsell, 1998; Toth & Assad, 2002), and the encoding of non-spatial information in LIP was task-dependent and only occurred when such information was

behaviorally relevant (Toth & Assad, 2002). In addition to basic visual features, task-relevant information encoding was also seen for higher level representations such as category membership and the associations between stimuli (Fitzgerald, Freedman, & Assad, 2011; Fitzgerald et al., 2013; Freedman & Assad, 2006; Swaminathan & Freedman, 2012). These findings suggest that LIP neurons do not just simply encode sensory stimuli, but can represent meaningful information such as category membership that is extracted from perceptual input.

The IPS region in humans shows similar functional properties to monkey LIP. For instance, using fMRI, it has been found that human IPS shows sensitivity to non-spatial information such as object shapes (Konen & Kastner, 2008) and encodes task-relevant visual information such as VSTM representations (Xu & Chun, 2006; Xu & Jeong, in press). The response amplitude in the superior part of IPS has been shown to correlate with VSTM capacity, suggesting this region encodes VSTM contents (Todd & Marois, 2004; 2005; Xu & Chun, 2006). The encoding of VSTM contents in superior IPS was also reported in a recent multi-voxel pattern analysis (MVPA) study (Xu & Jeong, in press). These findings suggest the role of IPS in the dorsal pathway is broader than the two-pathway view suggests.

0.3 Plan of dissertation

Based on previous findings that showed the encoding of visual object information in human IPS, in this dissertation, I further investigated whether flexible visual representation exists in human IPS region. To examine flexible visual information

processing in IPS, it is important to localize a region that is directly involved in visual object encoding. As such, superior IPS is a logical candidate region, as previous work has demonstrated that it is directly involved in the encoding of behaviorally relevant visual information (Todd & Marois, 2004; 2005; Xu, 2010; Xu & Chun, 2006; Xu & Jeong, in press).

In the current work, an independent VSTM task that varied set size was used to localize superior IPS. A multiple regression analysis was then performed with coefficients weighted by the VSTM capacity, as measured behaviorally, for each set size to identify regions with activity that parallels behavioral performance. This approach has consistently identified the superior part of IPS, in both hemispheres, suggesting this regions is involved in the representation of VSTM information (Todd & Marois, 2004; Xu & Chun, 2006).

In Chapter 1, I investigated how task-irrelevant object shape information is processed in superior IPS. I manipulated the numbers of target to be encoded in VSTM and distractors, and measured fMRI response amplitudes in superior IPS. I found increased fMRI response in superior IPS when target shapes appeared with distractors, but only under low target encoding load. Moreover, the presence of distractors did not increase fMRI response in superior IPS when target locations were cued in advance. In contrast, when I examined inferior IPS, a region along IPS involved in object individuation, and the lateral occipital (LO) region, which is involved in visual object processing, I found that these regions also encoded distractors under low target encoding load, regardless of whether target locations were cued or not.

In Chapter 2, using MVPA (Cox & Savoy, 2003; Haxby et al., 2001; Norman, Polyn, Detre, & Haxby, 2006), I examined whether superior IPS can extract different feature

information from the same object depending on the task demands. To do this, I kept the visual input the same across tasks, but asked participants to attend to different feature dimension(s) of the objects. I found that even with the same visual input, shape information was only decoded in superior IPS when it was task-relevant. In inferior IPS shape could be decoded regardless of task demands, but decoding accuracy was higher when shape was task relevant, suggesting some degree of task-dependent modulation. LO did not show any task-dependent representations.

In Chapter 3, I tested whether task-dependent representations in superior IPS are limited to basic-level visual features such as shape or can be extended to abstract information such as viewpoint invariant object identity. Across three experiments, I found that superior IPS can form identity representations that are extracted from perceptually distinct images (e.g., faces from different viewpoints, and with different hairstyles and facial expressions). Such abstract identity representation was not found in other ventral visual regions such as LO, fusiform face area (FFA), parahippocampal place area (PPA), and visual word form area (VWFA). Furthermore, I found that the neural representation of identity in superior IPS significantly correlates with a behavioral measure of identity similarity, confirming that the neural representation is reflecting the perceived identity information.

Taken together, the results of these three chapters show that human superior IPS flexibly encodes a variety of visual information, from simple shape to abstract identity. Furthermore, the current findings demonstrate that the representation in superior IPS is

dynamically modulated by task demands. These findings suggest that superior IPS plays a key role in mediating flexible visual information processing in the human brain.

1

Neural representation of targets and distractors during object individuation and identification

1.0 Abstract

In many everyday activities, we need to attend and encode multiple target objects among distractor objects. For example, when driving a car on a busy street, we need to simultaneously attend objects such as traffic signs, pedestrians, and other cars, while ignoring colorful and flashing objects in display windows. To explain how multiple visual objects are selected and encoded in visual short-term memory (VSTM) and in perception in general, the neural object file theory argues that whereas object selection and individuation is supported by inferior intra-parietal sulcus (IPS), the encoding of detailed object features that enables object identification is

mediated by superior IPS and higher visual areas such as the lateral occipital complex (LOC). Nevertheless, because task-irrelevant distractor objects were never present in previous studies, it is unclear how distractor objects would impact neural responses related to target object individuation and identification. To address this question, in two fMRI experiments, we asked participants to encode target object shapes among distractor object shapes, with targets and distractors shown in different spatial locations and in different colors. We found that distractor-related neural processing only occurred at low, but not at high, target encoding load and impacted both target individuation in inferior IPS and target identification in superior IPS and LOC. However, such distractor-related neural processing was short-lived as it was only present during the VSTM encoding but not the delay period. Moreover, with spatial cuing of target locations in advance, distractor processing was attenuated during target encoding in superior IPS. These results are consistent with the load-theory of visual information processing. They also show that while inferior IPS and LOC were automatically engaged in distractor processing under low task load, with the help of precuing, superior IPS was able to only encode the task-relevant visual information.

1.1 Introduction

Encoding, retaining, and retrieving visual information relevant to behavior and thoughts are some of the most fundamental human cognitive abilities. Over the past six decades, pioneered by human neuropsychological studies on patients such as H.M. (Corkin, 1968; B. Milner, Corkin, & Teuber, 1968; Scoville & Milner, 1957) (;see also Corkin, 2002; Corkin, Amaral,

González, Johnson, & Hyman, 1997), many insights have been gained regarding the role of the medial temporal lobe in mediating information retention in long-term memory. Meanwhile, how visual information is first perceived and retained in visual short-term memory (VSTM) has been linked to the functions of the prefrontal cortex and the parietal cortex (Goldman-Rakic, 1995; Todd & Marois, 2004; Ungerleider, Courtney, & Haxby, 1998; Xu & Chun, 2006).

In one study using functional magnetic resonance imaging (fMRI), Xu and Chun (2006) asked participants to encode multiple object shapes into VSTM. They found that responses in inferior intra-parietal sulcus (IPS) increased with increasing object number and plateaued at about set size 4 regardless of object complexity. In addition, they found that responses from superior IPS and lateral occipital complex (LOC, an object shape area, see Malach et al., 1995) increased with set size and plateaued at about the maximal number of objects held in VSTM (equal or less than four) as determined by object complexity. Based on these and other related findings, Xu and Chun proposed the neural object file theory and argued that, in VSTM as well as in perception in general, object individuation is supported by inferior IPS and object identification is mediated by superior IPS and higher object processing regions such as LOC (see also Xu, 2007; 2008; 2009; Xu & Chun, 2007; 2009). Here, object individuation refers to the selection of objects via their spatial locations, whereas object identification refers to the encoding of detailed object featural information. These neural findings are in line with previous behavioral findings and theories regarding how the visual system selects and encodes multiple objects through individuation and identification processes (Kahneman, Treisman, & Gibbs, 1992; Pylyshyn, 1989; 1994).

Nevertheless, because only targets were included in previous studies (Xu, 2007; 2009; Xu & Chun, 2006; 2009), it is unclear how the neural mechanisms mediating object individuation and identification would operate in the presence of task-irrelevant distractors. Understanding the impact of distractors during object individuation and identification is essential if we want to generalize laboratory findings to real world object perception, as irrelevant visual information is always present in everyday vision.

How distractors are filtered out by our visual system has been examined by research dated back to the 1950s. The early-selection view argues that the visual system can select targets and ignore distractors very early on during visual processing (Broadbent, 1958; Moray, 1959). According to this view, the presence of distractors should have minimum impact on the neural responses mediating visual object individuation and identification. Alternatively, the late-selection view argues that our visual system can individuate or even identify distractors (J. A. Deutsch & Deutsch, 1963; Luck et al., 1996). According to this view, the presence of distractors would significantly impact neural substrates supporting object individuation and identification. A third possibility is that the processing of distractors depends on the available resources. Accordingly, irrelevant information is processed only when the main task is relatively easy and does not consume all the available resources (Lavie, 2005; Lavie & Tsal, 1994; see also Yi et al., 2004). This view would predict that distractors will only be processed and impact neural responses when the demand for object individuation and identification is low. Indeed, when Xu (2010) examined the encoding of two features from the same object, with one being task-relevant and the other task-irrelevant, she found that object-based encoding of task-irrelevant object features only occurred when the demand to encode task- relevant object features was low.

Moreover, such object-based processing was short-lived and was not sustained over a long delay period.

In the present study, we investigated the impact of task-irrelevant distractors on the neural mechanisms supporting object individuation and identification when targets and distractors appeared in different spatial locations. In Experiment 1, we asked participants to encode target shapes among distractor shapes in a VSTM task. A long delay period was used to allow us to separately examine encoding-, delay-, and retrieval-associated neural responses. In Experiment 2, we asked whether top-down attention could modulate distractor processing during object individuation and identification. By using either neutral or valid location cues, we tested whether distractor processing could be excluded when participants knew target locations in advance.

1.2 Experiment 1

In this experiment, we examined the impact of distractors on object individuation and identification during both the VSTM encoding and delay periods. We varied the target load by presenting either 1 or 4 target shapes in one color, and varied the distractor load by presenting either 0 or 3 distractor shapes in a different color. We measured neural responses in independently defined inferior IPS, superior IPS, and LOC regions of interest (ROIs). The early selection theory would predict that distractors would be processed regardless of the encoding demand. The late selection theory, on the other hand, would predict that distractors would be filtered out during the encoding period. Lastly, the load theory would predict that processing of

distractors would depend on the target encoding load.

1.2.1 Methods

Participants

Twelve paid participants (7 females) were recruited from the Harvard University community (mean age 23.83, SD = 4.87) with informed consent, which was approved by the Institutional Review Board of Harvard University. All of them were right-handed and had normal or corrected to normal visual acuity. One additional participant was tested but was excluded from further analysis due to excessive head motion (more than 5 mm).

Main Experimental Design

The participants were asked to remember target shapes among distractor shapes presented briefly around the central fixation. After an extended delay, they judged whether a probed shape matched one of the remembered target shapes by pressing either the "match" or the "no-match" key (see Figure 1 for an illustration of the trial sequence). A match occurred in half of the trials. Targets and distractors were shown in different colors to facilitate target selection, with half the participants having red targets and green distractors and the other half having the reverse color assignment.

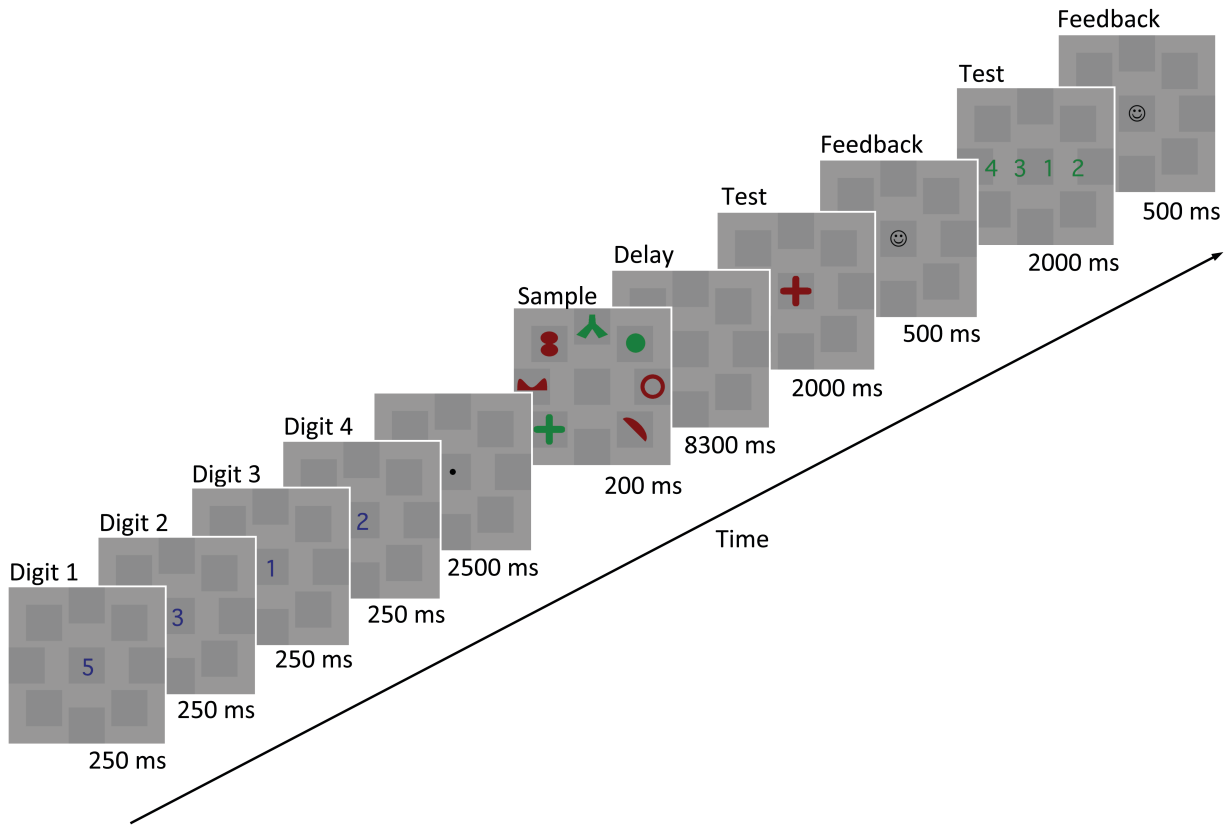

Figure 1. An example trial of Experiment 1. Participants were asked to remember target shapes (in red) among distractor shapes (in green). After an extended delay, they judged whether the probed shape matched one of the remembered target shapes by pressing the appropriate response button. A match occurred in half of the trials. Target and distractor color assignment was balanced across different participants. To discourage grouping, eight dark grey squares were also present as placeholders and marked all possible locations that targets and distractors could appear. To prevent verbal encoding of the shapes, in addition to the VSTM shape task, 4 digits were shown sequentially at the beginning of each trial. Participants were asked to remember and rehearse these digits and then judge whether the same 4 digits were shown at the end of the trial.

There were 4 conditions: 1 target with 0 distractors (1T), 1 target with 3 distractors (1T+3D), 4 targets with 0 distractors (4T), and 4 targets with 3 distractors (4T+3D). All stimuli appeared on a light grey background. To prevent grouping, eight dark grey squares were also presented as placeholders and marked all the possible locations for which targets and distractors could appear (see Figure 1, see also, Xu, 2009). Eight different target and distractor shapes were used (see Xu & Chun, 2006), each subtended approximately 2.74° x 2.74°. The size of the entire

display was 11.8° x 11.8°.

To prevent participants from verbally encoding the shapes, in addition to the VSTM shape task, they were required to remember and rehearse four digits throughout each VSTM trial by comparing whether four digits presented sequentially at the beginning of each trial matched those presented simultaneously at the end of each trial. Inferior IPS has been shown to track the number of objects presented at different spatial locations (up to 4 locations, see Xu, 2009; Xu & Chun, 2006). As such, given the 6 sec lag in hemodynamic response, simultaneous presentation of the four digits at different spatial locations may saturate inferior IPS response before the presentation of the target and distractor stimuli (which occurred 2.5 sec after the digit presentation). For this reason, digits were presented sequentially, rather than simultaneously, at the beginning of each trial. Each trial lasted 18 sec and contained the followings: a fixation period (1000 ms), a sequential presentation of four digits (250 ms each), a fixation period (2500 ms), a sample shape display (200 ms), a delay period (8300 ms), a test shape display (2000 ms), a shape response feedback (500 ms), a test digit display (2000 ms), and a digit response feedback (500 ms) (Figure 1). The participants were instructed to maintain fixation during the trial. With a counterbalanced trial history design (see Todd & Marois, 2004; Xu & Chun, 2006), each run contained a total of 27 trials, including 5 trials for each stimulus condition, 5 fixation trials, and 2 filler trials. Fixation trials contained the digit task without the VSTM shape task (the shape task was replaced by a fixation dot). Filler trials were included to balance trial history, with one appearing at the beginning and one at the end of the run. Filler trials were excluded during data analysis. Each participant completed 4 or 5 runs, with each run lasting 8 min and 15 sec.

Localizer Design

To ensure that the ROIs we localized were involved in processing the specific visual stimuli used in the main experiment, the shapes from the main experiment appeared in the same size and eccentricity in all the ROI localizers described below as they did in the main experiment.

To localize the superior IPS region that closely tracks the amount of visual information retained in VSTM, we conducted an independent shape VSTM experiment similar to that of Xu and Chun (2006). Specifically, participants were asked to remember 1, 2, 3, 4, or 6 black object shapes presented briefly around the central fixation. After a short delay, a probe shape appeared at fixation and required participants to make a probe match/no-match judgment. The probe matched one of the remembered shapes in half of the trials. Each trial lasted 6 sec and contained the followings: a fixation period (1000 ms), a sample display (200 ms), a delay period (1000 ms), a test shape display/response period (2500 ms), and a feedback (1300 ms). The sizes of the individual object shape and the whole display were identical to those used in the main VSTM experiment. With a counterbalanced trial history design, there were 12 stimulus trials for each set size condition as well as 12 fixation trials in which only a fixation dot appeared during the 6-sec trial period. Three filler trials were added to the beginning and one filler trial was added to the end of each run for practice and trial history balancing purposes. These filler trials were excluded during data analysis. Each participant was tested with 3 runs, each lasting 7 min 42 sec.

To define the LOC and the inferior IPS ROIs, the same localizer experiment used in Xu and Chun (2006) was conducted here. Participants viewed blocks of object and noise images (both subtended 11.8° x 11.8°). The object images were the set size 6 displays used in the superior IPS localizer experiment. Each block lasted 16 sec and contained 20 images, with each image appearing for 500 ms and followed by a 300 ms blank delay. To engage participants' attention to

the displays, they were asked to detect a slight spatial jitter which occurred randomly once in every ten images. Each run contained 8 object blocks and 8 noise image blocks. Each participant was tested with 2 runs, each lasting 4 min and 40 sec.

fMRI methods

fMRI data were acquired from a Siemens Tim Trio 3T scanner at the Harvard Center for Brain Science in Cambridge, MA. Participants viewed images back projected onto a screen at the rear of the scanner bore through an angled mirror mounted on the head-coil. All experiments were controlled by an Apple MacBook Pro running Matlab with Psychtoolbox extensions (Brainard, 1997). Anatomical images were acquired using standard protocols. For both the localizer runs and the main experimental runs, 24 5-mm-thick (3 mm x 3 mm in plane, 0 mm skip) slices parallel to the AC-PC line were acquired using a gradient echo pulse sequence (TE 25ms, flip angle 90°, matrix 64 x 64). In the main VSTM experiment and the superior IPS localizer runs, TR of 1.5 sec was used; and in the inferior IPS localizer runs, TR of 2.0 sec was used.

Data analysis

Behavioral VSTM capacity for each set size was measured using Cowan's K formula which estimates the number of items retained in VSTM while controlling for correct guesses (K = (hit rate + correct rejection rate – 1) x N, where K is the number of items encoded in VSTM and N is the set size, see Cowan, 2001 for details).

fMRI data were analyzed with BrainVoyager QX 2.1 (www.brainvoyager.com). 3D

motion correction, slice acquisition time correction, linear trend removal, and Talairach space transformation were conducted during data pre-processing (Talairach & Tournoux, 1988).

To define the superior IPS ROI in each participant, as was done previously (Todd & Marois, 2004; Xu & Chun, 2006), fMRI data from the superior IPS localizer experiment were analyzed using multiple regressions with the regression coefficient for each VSTM set size weighted by that participant's behavioral VSTM capacity for that set size. The superior IPS was defined as voxels showing a significant activation in the regression analysis (false discovery rate $q < .05$, corrected for serial correlation) and whose Talairach coordinates matched those reported in Todd and Marois (2004). The LOC and inferior IPS ROIs were defined as voxels showing higher activations to the shape than to the noise displays (false discovery rate $q < .05$, corrected for serial correlation) in lateral occipital cortex and IPS respectively. Example superior IPS, inferior IPS, and LOC ROIs are shown in Figure 2.

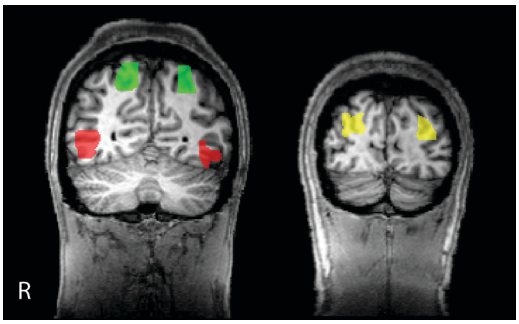

Figure 2. The superior IPS (green), the inferior IPS (yellow), and the LOC (red) ROIs from one example participant.

To examine responses from the main experiment, time courses from each participant in the main experiment were extracted from the three ROIs defined above. These time courses were

converted to percent signal change for each stimulus condition by subtracting the corresponding value for the fixation trials and then dividing by that value (see Kourtzi & Kanwisher, 2000; Todd & Marois, 2004; Xu & Chun, 2006). To capture VSTM encoding-related peak responses in each participant and to account for temporal variability of fMRI peak responses among the different participants, VSTM encoding-related peak responses from all participants were aligned to the 9th sec (6th TR) from the start of the trial. This anchor point was chosen based on responses from the majority of the participants. Time course data either remained the same, or was shifted forward or backward by 1.5 sec (1 TR) during this alignment process. To ensure that baseline fMRI response differences before the onset of the VSTM shape display would not contribute to peak fMRI response amplitude estimates, we calculated baseline response drift by averaging the responses from the first 6 seconds of each trial and then subtracted this drift from each point of the time course. This was done separately for each participant for each stimulus condition of each ROI.

1.2.2 Results

Behavioral results

The capacity of VSTM was estimated using Cowan's K formula (Cowan, 2001). The mean K values for the four stimulus conditions were 0.89±0.04 (1T), 0.90±0.04 (1T+3D), 1.90±0.35 (4T), and 1.79±0.3 (4T+3D). A two-way repeated measures ANOVA with target number (1 vs 4) and distractor number (0 vs 3) revealed a main effect of target number, $F(1,11) = 11.063$, $p = .007$, showing that more information could be retained in VSTM from 4 than from 1 target. No other

main effects or interactions reached significance ($F_s < 1$, $p_s > .57$).

Response times for the four stimulus conditions were 823±49 ms (1T), 828±36 ms (1T+3D), 1003±47 ms (4T), and 965±45 ms (4T+3D) respectively. Similar to the K measures, a two-way repeated measures ANOVA with target number and distractor number revealed a main effect of target number, $F(1,11) = 37.31$, $p < .001$, and a marginally significant interaction between target number and distractor number, $F(1,11) = 3.4$, $p = .092$. No other main effect reached significance ($F < 1$, $p > .37$).

fMRI results

fMRI responses from the main VSTM task were extracted from independently localized LOC, inferior IPS, and superior IPS ROIs. Percent signal change compared to fixation was calculated for each time point and the final time courses were plotted in Figure 3. These time courses showed two peaks, corresponding to the encoding of the initial shape display and the shape probe, respectively.

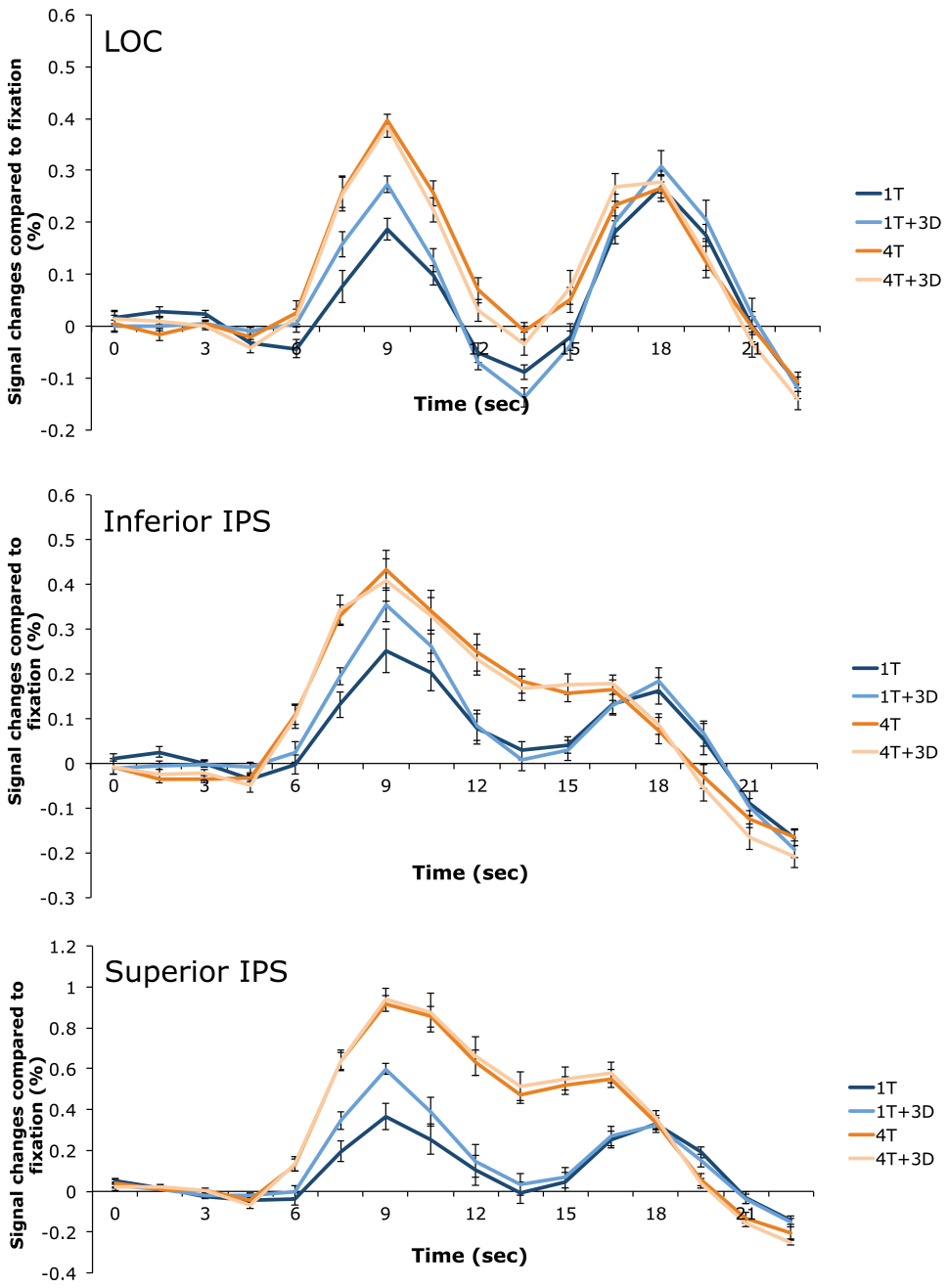

Figure 3. fMRI responses from LOC (top), inferior IPS (middle), and superior IPS (bottom) in Experiment 1. All three brain regions showed a similar response pattern. The presence of distractors only increased fMRI responses during the encoding period when the target encoding load was low. No distractor effect was present during the delay period. Blue line – one target; light blue line – one target and three distractors; orange line – four targets; and light orange line – four targets and three distractors. Error bars indicate within-subjects standard errors.

VSTM encoding-related activities

To examine VSTM encoding related activities, we analyzed the first fMRI peak responses at the 9th sec (6th TR) in the three ROIs. The effect of target number was present in superior IPS and LOC (Fs > 25.47, ps < .001), but not in inferior IPS (F < 1.93, p > .19). The effect of distractor number was present in superior IPS ($F(1,11) = 10.001$, $p = .009$), but not in the other two brain regions (Fs < 2.41, ps > .148). Importantly, all three brain regions showed a significant interaction between target number and distractor number (Fs > 13.27, ps < 0.004), indicating that distractor encoding was greater when one target than when four targets had to be encoded. Confirming this last result, in pairwise comparisons, in all three ROIs, significant differences were observed between 1T and 1T+3D conditions (Fs > 3, ps < 0.05), but not between 4T and 4T+3D conditions (Fs < 1, ps > 0.58). These results showed that distractor processing in inferior and superior IPS depended on target encoding load and only occurred at the low task load. Given that inferior and superior IPS have been proposed to be involved in object individuation and identification respectively (Xu, 2007, 2009; Xu & Chun, 2006, 2009), these results suggest that distractor processing impacts both stages of object processing and is load dependent.

We also compared the difference between the 1T+3D and 4T conditions in which the same total number of items were presented but target number differed. Interestingly, the difference between these two conditions was not significant in inferior IPS ($F < 1$, $p = 0.37$), but reached significance in both superior IPS and LOC (Fs > 4.67, ps < .01). In fact, the difference between these two conditions was greater in superior IPS than in inferior IPS ($F = 2.33$, $p = .039$). This may explain why we failed to obtain a main effect of target number in inferior IPS.

These results indicate that, when distractors were encoded under low target load, they

were not differentiated from targets in inferior IPS that supports object individuation; the difference between targets and distractors only emerged in superior IPS and LOC that support object identification. This is consistent with the predictions of the neural object file theory proposed by Xu and Chun (2009). They argued that only object location information is predominantly encoded during object individuation and that detailed object feature information becomes available later during object identification-related processing (see also, Xu, 2009).

VSTM delay-related activities

To examine VSTM maintenance related activities, we analyzed fMRI responses at the 13.5th sec (9th TR) when responses reached a minimum before they started to rise again with the presentation of the probe display. During this delay period, a main effect of target number was observed in all three ROIs ($Fs > 10.24$, $ps < 0.01$), showing that four target conditions elicited higher responses than one target conditions. A main effect of distractor number was observed in LOC ($F(1,11) = 6.796$, $p = .024$), showing a lower response for distractor present than for distractor absent conditions. Critically, there was no interaction between target number and distractor number in all three ROIs ($Fs < 1$, $ps > .5$). These results indicated that distractors either had no impact on target processing, or they were completely suppressed during the delay period. Either way, distractor processing did not depend on target processing load.

1.2.3 Discussion

By examining the impact of distractors on object individuation and identification during

VSTM encoding and delay periods, here we observed neural encoding of distractors during both object individuation and identification when the target encoding load was low. The encoding of distractors under low load is consistent with the predictions of the load theory (Lavie, 2005; Lavie & Tsal, 1994).

Such load-dependent distractor response in inferior and superior IPS and LOC distinguishes them from pure stimulus-driven retinotopic visual regions. This is because, while almost twice the area was stimulated when four targets were presented with three distractors than when they were presented alone, we failed to observe any increase in response amplitude in these three brain regions.

When Xu (2010) examined the encoding to two features from the same object, she found that object-based encoding of task-irrelevant distractor features only occurred when the demand to encode the task-relevant target features was low. Because target and distractor features appeared on the same object and at the same location in Xu (2010), it might have been difficult to suppress the processing of distractor features. However, the present experiment showed that, even when targets and distractors appeared in different spatial locations and in different colors, distractor processing still could not be suppressed at low task encoding load. This indicates that the encoding of distractors at low task load may be automatic and obligatory.

Meanwhile, the present experiment showed that the neural response for distractors was short lasting and quickly decayed when no attempt was made to sustain it during the subsequent delay period. This is consistent with Xu (2010) which showed a similar response pattern for task-irrelevant features during object-based feature encoding. Thus, although the neural encoding of distractors at low target load may be initially automatic and obligatory, participants can exert

control over what is retained for a prolonged period of time.

1.3 Experiment 2

In the Posner cueing paradigm (Posner, 1980), participants can better detect targets present at the cued than at the uncued spatial locations. This shows that the deployment of spatial attention can prioritize the processing of visual information at specific locations. Can such top-down attentional control suppress the processing of task-irrelevant distractors during target object individuation and identification? It is possible that with spatial cuing, neural encoding of distractors at low task load can be completely suppressed. It is equally likely, however, that while the processing of distractors is attenuated, it cannot be completely suppressed, and that different amount of suppression may occur during target object individuation and identification. In this experiment, to understand how automatic and obligatory it is to encode task-irrelevant distractors under low load, we precued the locations of the targets before target onset and tested whether distractor encoding could be suppressed by top-down attention. Given that Experiment 1 showed that the presence of distractors had no impact on VSTM maintenance and retrieval related activities (see Figure 3), to streamline our design, instead of using a 8.3 sec delay period, here we used a 1 sec delay period.

1.3.1 Methods

Participants

Nine new participants (seven females) were recruited from the Harvard University

community (mean age 28.33, SD = 4.52) with informed consent, which was approved by the Institutional Review Board of Harvard University. All of them were right-handed and had normal or corrected to normal visual acuity. One additional participant was tested but excluded from further analyses due to excessive amount of head motion.

Design

The main VSTM experiment was identical to Experiment 1 except for the followings. We shortened the delay period to 1000 ms, as the focus of this experiment was on distractor encoding. We also removed the verbal rehearsal load, as VSTM task performance with a short delay period has been shown to be unaffected whether verbal rehearsal is imposed or not (Luck & Vogel, 1997). In the valid-cue trials, we cued target locations by rapidly flashing small dots twice at the target location prior to target onset. The neutral-cue trials were similar to the valid-cue trials, except that all 8 locations where targets and distractors could possibly appear were cued by the flashing dots. To maximize the effect of cuing, valid- and neutral-cue trials were shown in different runs, with half of the participants tested with the valid-cue trials before the neutral-cue trials and the other half had the reverse order of testing. The exact timing of a trial was as follows: first precue (125 ms), a fixation period (125 ms), second precue (125 ms), a fixation period (625 ms), a sample display (200 ms), a delay period (1000 ms), a test shape display (1800 ms), and a feedback (2000 ms). Note that the 1000 ms interval between the initial onset of the cue and the onset of the stimulus display was the same as that used in Posner (1980). The participants were instructed to maintain fixation at the center fixation dot and covertly pay attention to the cued locations. Other aspects of this experiment were identical to those of Experiment 1.

Data analyses

Because each trial lasted 6 sec with a 1 sec delay period, only one fMRI response peak was observed, reflecting the summed fMRI responses from VSTM encoding, maintenance and retrieval. As such, instead of presenting data from each time point as we did in Experiment 1, only peak responses were extracted and included in further statistical analyses. All other aspects of data analyses were identical to that of Experiment 1.

1.3.2 Results

Behavioral results

K values were 0.97±0.015 (1T), 0.98±0.01 (1T+3D), 3.2±0.16 (4T), and 2.71±0.27 (4T+3D) for neutral-cue trials, and 0.97±0.015 (1T), 0.97+0.015 (1T+3D), 2.9±0.15 (4T), and 2.93±0.17 (4T+3D) for valid-cue trials. A three-way repeated measures ANOVA with cue type (neutral vs valid), target number (1 vs 4), and distractor number (0 vs 3) was conducted. Main effect of target number was significant, $F(1,8) = 220.44$, $p < .001$, showing that more information was stored in VSTM when target number was 4 than 1. No other main effects or interactions reached significance (ps > .16).

Response times were 496.6±31.8 ms (1T), 481.4±28.4 ms (1T+3D), 753.6±48.2 ms (4T), and 760.5±44.2 ms (4T+3D) for neutral-cue trials, and 520.7 ms±26.7 (1T), 501.6±24.9 ms (1T+3D), 763±48.4 ms (4T), and 776.1±44.1 ms (4T+3D) for valid-cue trials. A three-way ANOVA with cue type, target number, and distractor number revealed a main effect of target

number ($F(1,8) = 131.08$, $p < .001$), showing that response time was slower when more targets had to be encoded and retrieved for comparison. There was also an interaction between target number and distractor number ($F(1,8) = 5.521$, $p = .047$), indicating that response time difference between 1 and 4 target trials were larger when there were 3 than 0 distractor. This is likely associated with the greater effort needed to filter out distractors at high than at low target encoding load. No other main effects or interactions reached significance (ps > .36).

fMRI results

Although the peak fMRI responses examined here reflected the summed fMRI responses from VSTM encoding, maintenance and retrieval periods, given that Experiment 1 showed that the presence of distractors had no impact on maintenance and retrieval related activities (see Figure 3), any distractor effect we obtained here could only come from encoding related activities.

In all three ROIs, as can be seen in Figure 4, there were a main effect of targets, a main effect of distractors, and an interaction between the two (all Fs > 9.24, ps < 0.05). This replicated our findings from Experiment 1 and showed that the presence of distractors significantly impacted target processing in a load-dependent manner.

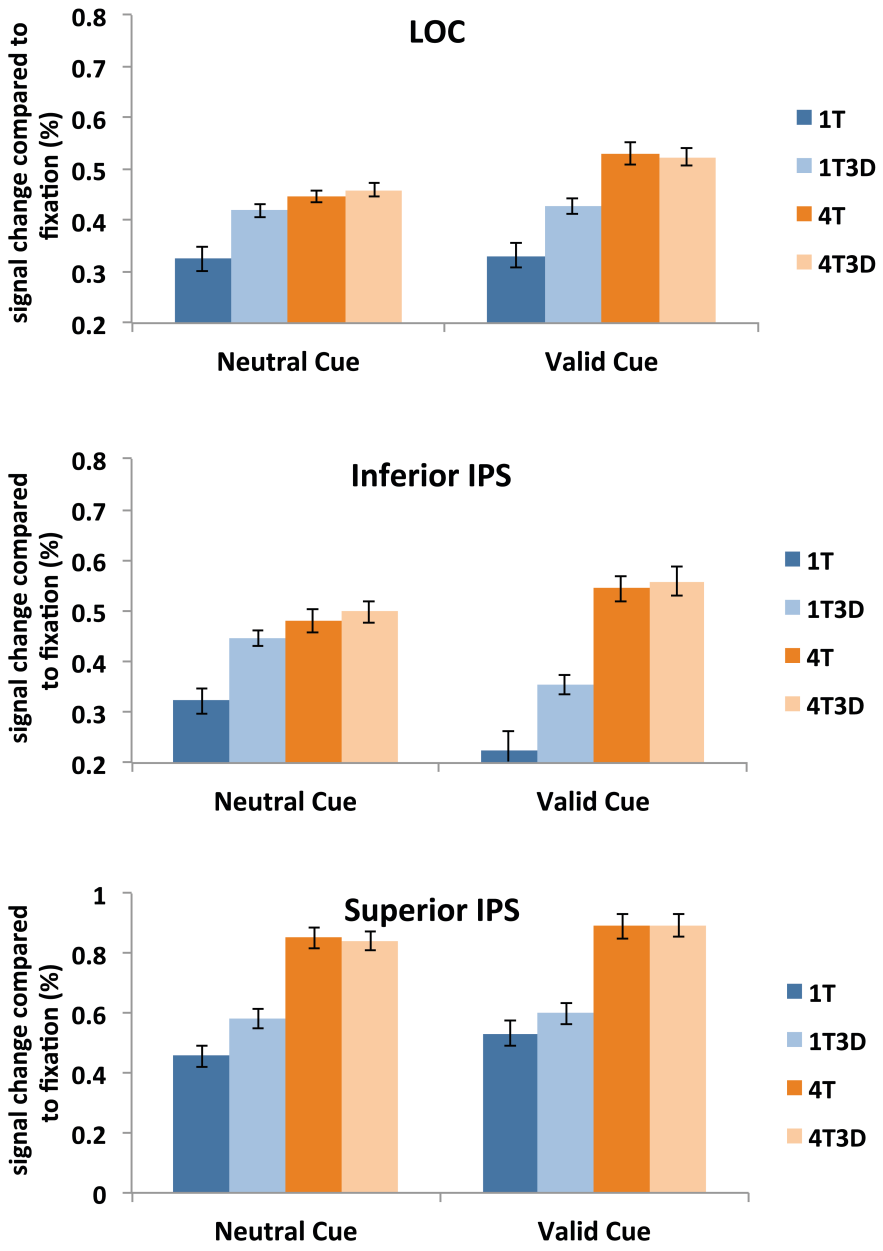

Figure 4. fMRI responses from LOC (top), inferior IPS (middle), and superior IPS (bottom) in Experiment 2. As in Experiment 1, the presence of distractors only increased fMRI responses when the target encoding load was low. Critically, although spatial cuing did not completely remove distractor processing, it did significantly attenuate distractor processing during target object processing in superior IPS. Blue bar – one target; light blue bar – one target and three distractors; orange bar – four targets; and light orange bar – four targets and three distractors. Error bars indicate within-subjects standard errors.

Of main interest was the effect of spatial cuing. Out of the three ROIs, only the superior

IPS showed a significant 3-way interaction of cue type, target number and distractor number ($F(1,8) = 5.827$, $p = .042$; for inferior IPS and LOC, $Fs < 1$, $ps > .6$). Detailed comparisons revealed that, in superior IPS, under low target encoding load, although the effect of distractor was still present in both the valid cue and the neutral cue conditions ($F = 2.74$, $p = .025$; and $F = 5.77$, $p < .001$, respectively), distractor processing was significantly attenuated with spatial cuing, resulting in a significant interaction between cue type and distractor number in the 1 target conditions ($F(1,7) = 18.167$, $p = .003$). Such an interaction, however, was absent in the 4 target conditions ($F < 1$, $p > .63$). This pattern of response was found in every single one of our participants. Comparing directly across the three ROIs, there was a marginally significant interaction between the effect of cuing under low load and brain region ($F(2, 16) = 3.27$, $p = .064$), showing that the effect of cuing under low load was stronger in superior IPS than in the other two brain regions.

In inferior IPS and LOC, there was an interaction between cue type and target number ($Fs > 16.8$, $ps < .01$), showing that the difference between 1 and 4 target conditions was greater in the valid than in the neutral cue conditions. This could be due to differences in cue-related encoding, as in the valid-cue trials, 1 cue and 4 cues were shown for the 1 and 4 target conditions, respectively; whereas in the neutral-cue trials, 8 cues were always shown regardless of the target encoding load. It is also possible that this interaction between cue type and target number was the result of more efficient allocation of resources with target cuing, such that less resources were allocated to the 1 target conditions in the valid than in the neutral cue conditions, and more resources were allocated to the 4 target conditions in the valid than in the neutral cue conditions.

1.3.3 Discussion

This experiment replicated the main findings of Experiment 1 and showed that distractor processing only occurred under low target encoding load. Importantly, this experiment indicated that although spatial cuing could not completely remove distractor processing, it could significantly attenuate distractor processing in superior IPS. This is consistent with a previous finding showing that superior IPS is mainly involved in processing what is most task-relevant (Xu, 2010).

Unlike Posner (1980), here we did not observe any behavioral cuing benefit. This is likely due to the fact that our VSTM paradigm is not configured to produce the behavioral cuing benefit. In Posner's study, participants made speeded detection for the appearance of the cued target. In the present experiment, this was not measured. Rather, behavioral accuracy and RT mainly reflected responses to the shape probe one second after the presentation of the target shapes. Nevertheless, the effect of cuing did impact distractor processing in superior IPS, showing that in this case fMRI measures could be more sensitive and informative than behavioral measures.

In our experiment, we used a fixed time interval between the initial onset of the cue and the onset of the stimulus display. It would be worth manipulating this interval in future studies to see whether distractor processing is modulated by this interval during target object individuation and identification. In any event, the 1000 ms cuing interval used in this experiment, which was

the same as that used in Posner (1980), clearly illustrates the feasibility of using spatial cuing to prioritize the processing of targets among distractors.

1.4 General Discussion

In this study, we investigated the processing of task-irrelevant information during visual object individuation and identification by examining the neural substrates mediating these processes. We asked participants to encode in VSTM target object shapes among distractor object shapes appearing at different spatial locations and in different colors and examined fMRI responses from parietal and occipital regions. In Experiment 1, we found that distractor processing depended on the availability of processing resources. Specifically, only when the demand to encode target shapes was low, did the presence of distractors increase neural responses in inferior IPS, LOC, and superior IPS. Given the involvement of these brain regions in object individuation and identification (e.g., Xu & Chun, 2009), these results suggest that, under low target encoding load, distractors were individuated and encoded. However, neural responses for distractors were short-lived as they were only present during the VSTM encoding period but not during the subsequent VSTM delay period. In Experiment 2, we examined whether distractor encoding under low task load could be suppressed if spatial attention was deployed ahead of the time to the target locations. Precuing target locations decreased distractor processing under low task load in superior IPS but not in inferior IPS or LOC. Thus, although distractor processing under low task load is obligatory and automatic during object individuation in inferior IPS and object encoding in LOC, it can be attenuated during object encoding in superior IPS with

precuing.

Consistent with this result, Xu (2010) reported that superior IPS encoded only task-relevant features regardless of the target encoding load whereas LOC encoded task-irrelevant information at low load. Likewise, task-dependent encoding in parietal regions has also been reported in neurophysiology studies (Freedman & Assad, 2006; Toth & Assad, 2002). Thus, although distractor processing was not suppressed in superior IPS in Experiment 1, with the help of precuing, this brain region can exhibit some degrees of task-dependent responses in Experiment 2.

Results of this study, together with previous studies showing the impact of perceptual and working memory load on distractor processing in other visual tasks, support the perceptual load theory which argues that the processing of distractors depends on the available resources and only occurs when the main task is relatively easy and does not consume all the available resources (Lavie, 2005; Lavie et al., 2004; Lavie & Tsal, 1994; Pinsk, Doniger, & Kastner, 2004; Torralbo & Beck, 2008; Xu, 2010; Yi et al., 2004). Meanwhile, the present study also identifies situations in which distractor processing under low task load may be suppressed (i.e., during the VSTM delay period) or substantially attenuated (i.e., with spatial cuing during object encoding in superior IPS).

Because distractor suppression related neuronal activities could also increase fMRI responses, one may argue that an increased fMRI response at low task load could reflect distractor suppression, rather than encoding. This, however, is unlikely the case due to the following two reasons. First, although distractor suppression was more critical at high task load when participants needed to dedicate all their encoding resources to targets, we did not see an

increased fMRI response to distractor processing at high task load in both Experiment 1 and 2. Second, with spatial cuing in Experiment 2, more suppression would be applied to distractors, and yet we observed an attenuated fMRI response to distractor processing at low task load in superior IPS and no response to distractor processing at high task load in all three ROIs. Thus, the distractor-related fMRI responses reported here reflect distractor encoding and not suppression.

Although the present study showed that distractors could be individuated and encoded when the target encoding load was low, it is unknown whether target and distractor shapes were encoded with the same precision. When they are task-relevant, shapes need to be encoded in sufficient resolution to support later memory recognition; when they are task-irrelevant, however, shapes may not be encoded in such fine resolution. Recent studies using multi-voxel pattern analysis (MVPA) have been able to decode visual information representation in a brain region by examining fMRI voxel response patterns (Cox & Savoy, 2003; Haxby et al., 2001; Norman et al., 2006). Further research using the MVPA approach may inform us of the exact nature of distractor shape representation during visual object individuation and identification.

In summary, the current study showed that, under low target encoding load, distractors elicited significant neural responses across a number of brain regions previously shown to be involved in visual object individuation and identification. This suggests that distractors are individuated and encoded at load target encoding load. However, such neural responses for distractors were short-lived as they were only present during the VSTM encoding but not the delay period. Although distractor processing was obligatory and automatic at low task load, with spatial cuing, it could be attenuated during object encoding in superior IPS.

2

Flexible visual information representation in human intra-parietal sulcus

2.0 Abstract

Even with identical visual input, our conscious mind can selectively encode what is most relevant to the current behavioral goals or thoughts of the observer. This mental ability requires the support of a neural mechanism that can flexibly encode a variety of visual information. In Macaque monkey electrophysiology studies, neurons in lateral intra-parietal sulcus (LIP) have been shown to exhibit such task-dependent encoding flexibility (Fitzgerald et al., 2011; Freedman & Assad, 2006; Toth & Assad, 2002). Here, we show that a similar neural mechanism exists in

human parietal cortex. Using functional magnetic resonance imaging (fMRI) and multi-voxel pattern analysis (MVPA), we found that, with identical visual input under different task conditions, neural response pattern in human superior intra-parietal sulcus (IPS) could decode object shape information only when it was required by the task. Inferior IPS (an adjacent region along IPS) carried object shape representation regardless of task demands, but the representation was modulated by task demands. However, other visual object area, such as lateral occipital area (LO), did not show task-dependent modulation of shape. These results show that human parietal regions are directly involved in object shape representation in a flexible manner. This capability likely places IPS as a key neural mechanism mediating the moment-to-moment visual information processing in the human brain.

2.1 Introduction

In everyday life, we often encounter multiple complex objects at the same time. To process such a huge amount of incoming visual information efficiently, it is important for our visual system to select behaviorally relevant information. Neurophysiological research has provided strong evidence of such flexible visual information processing in lateral intra-parietal (LIP) neurons and showed that these neurons encode behaviorally relevant visual stimuli (Gottlieb, Kusunoki, & Goldberg, 1998; Toth & Assad, 2002). Neuroimaging studies suggest similar task-relevant visual information processing occurs in human parietal cortex. For example, visual short-term memory (VSTM) (Baddeley, 1986; Luck & Vogel, 1997; Phillips, 1974), which stores task-relevant information, involves the recruitment of posterior parietal cortex. ERP

studies found that neural signal from posterior parietal region tracked the number of items held in VSTM (Vogel & Machizawa, 2004; Vogel, McCollough, & Machizawa, 2005). fMRI studies further identified intra-parietal sulcus (IPS), a sub region in parietal cortex, closely tracked VSTM capacity (Todd & Marois, 2004; 2005; Xu & Chun, 2006). Moreover, recent multi-voxel pattern analysis (MVPA) studies also support that posterior parietal region including IPS directly represent VSTM contents (Christophel & Haynes, 2014; Christophel, Hebart, & Haynes, 2012; but see Riggall & Postle, 2012). These findings suggest that IPS region encodes task-relevant visual information.

Nevertheless, it is not yet clear whether or not task-relevant visual information is flexibly represented in IPS. Specifically, it is still not known whether IPS region encodes all the features of task-relevant objects or extracts only task-relevant feature(s) from objects. To answer this question, in the current study, we investigated the neural representation of visual objects in IPS under four different task demands.

We examined three brain regions that are involved in visual object processing. Superior and inferior IPS are sub-regions along IPS previously shown to be involved in visual object identification and individuation, respectively (Xu & Chun, 2006; 2009). Superior IPS is previously shown to participate in object information encoding and storage in VSTM tasks (Todd & Marois, 2004; Xu & Chun, 2006; 2009). Therefore, it is a region where task-relevant visual information may be represented. Inferior IPS is anatomically close to superior IPS, but its functional role differs from superior IPS. Inferior IPS has previously been shown to participate in object selection and individuation via location and may contain coarse object information necessary for carrying out these operations (Jeong & Xu, 2013; Xu & Chun, 2006; 2009). In

addition, as a control region, we examined LO, which has been shown to be involved in visual object perception and recognition (Grill-Spector, Kushnir, Hendler, & Malach, 2000; Kourtzi & Kanwisher, 2000; Malach et al., 1995; A. D. Milner et al., 1991).

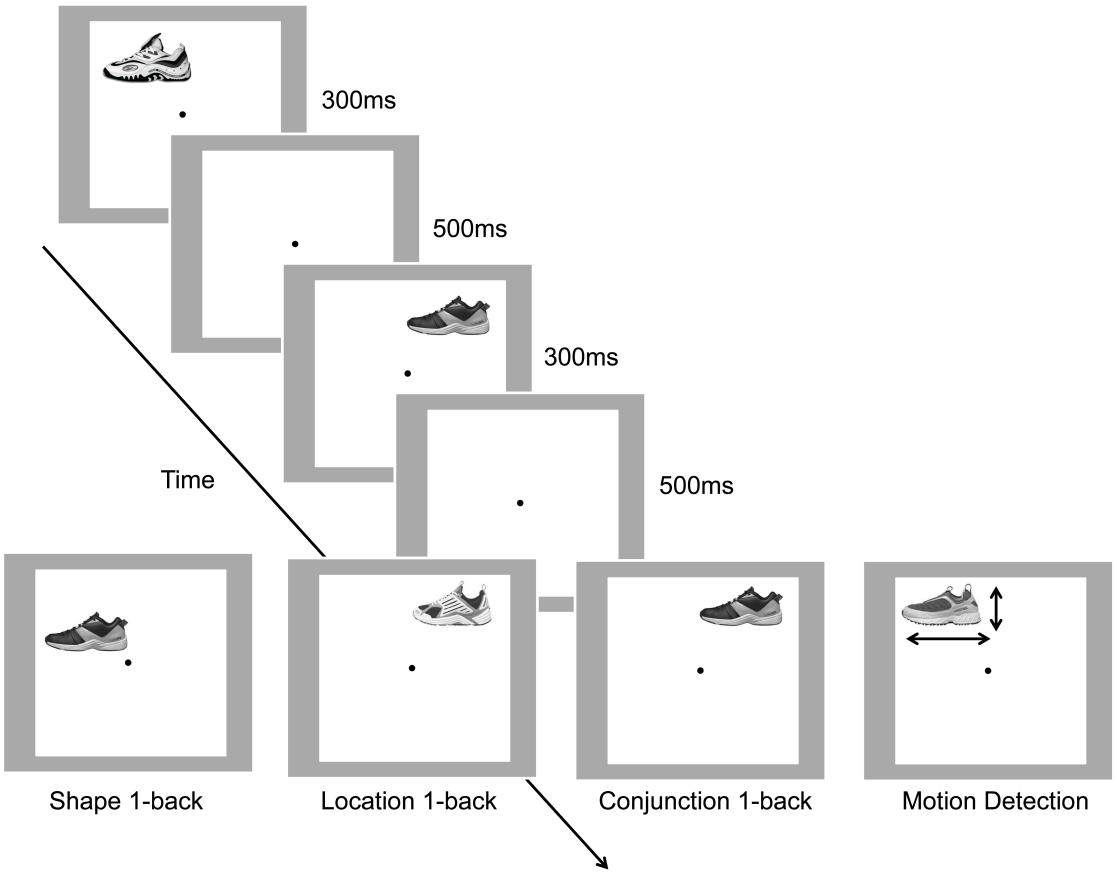

Figure 5. Illustration of the task. Participants saw the sequential presentation of images of one object category (e.g., shoe). In the shape, location, and conjunction 1-back tasks, participants monitored an immediate repetition of shape, location, or shape and location conjunction of the stimuli, respectively. In the motion detection task, they detected occasional vertical or horizontal movement of the stimuli (depicted as black arrows in the figure). The displays in the bottom depicted trials in which a repetition (in the 1-back tasks) or a motion (in the motion detection task) occurred.

While undergoing an fMRI scan, participants viewed the sequential presentation of images. They performed 4 different tasks in different runs within the same scan session: three 1-

back tasks requiring participants to detect an immediate repetition in either object shape, location or the conjunction of the two features, and a motion detection task. In each block of trials, ten object images from the same category that shared a general shape contour (e.g., 10 side-view shoe images, see Figure 5) were presented. The object images appeared randomly in one of four possible positions either all above or all below the central fixation dot. This allowed us to manipulate image location repetition between successively presented trials (see Methods).

Each task had identical visual input, but required different feature(s) to be task-relevant. In the shape 1-back task, participants viewed the images and detected an immediate repetition of the same exemplar while ignoring location changes, making only object shape task-relevant. In the location and conjunction 1-back tasks, they detected an immediate repetition of position of the exemplar and both position and identity of the exemplar, respectively. Thus, location but not shape was task-relevant in the location 1-back task, and both location and shape were task-relevant in the conjunction 1-back task. In the motion detection task, participants passively viewed the images and detected the direction of an occasional image jitter (either horizontal or vertical), making neither object shape nor location task-relevant.

2.2 Results

Participants' behavioral performance was fairly accurate across the four tasks (Mean±SD: 90.03±5.56%, 88.76±6.99%, 89.4±5.51%, and 93.15±5.35%, respectively, for the shape, location, conjunction 1-back tasks, and the motion task). There was a main effect of task, $F_{(3, 39)} = 3.33$, p

= .029, but differences between tasks did not reach significance in post-hoc pairwise comparisons (ps > .149, Bonferroni corrected).

Task-dependent encoding

To examine whether visual representation is modulated by task demands, we evaluated the decoding of task-relevant and –irrelevant feature dimensions in superior IPS, inferior IPS, and LO using a linear support vector machine (Cox & Savoy, 2003; Kamitani & Tong, 2005). For task-relevant decoding, we combined decoding accuracies for shape in the shape and conjunction 1-back tasks, and those for location in the location and conjunction 1-back tasks. For task-irrelevant decoding, decoding accuracies for shape in the location 1-back and motion tasks, and those for location in the shape 1-back and motion tasks were combined (see Methods).

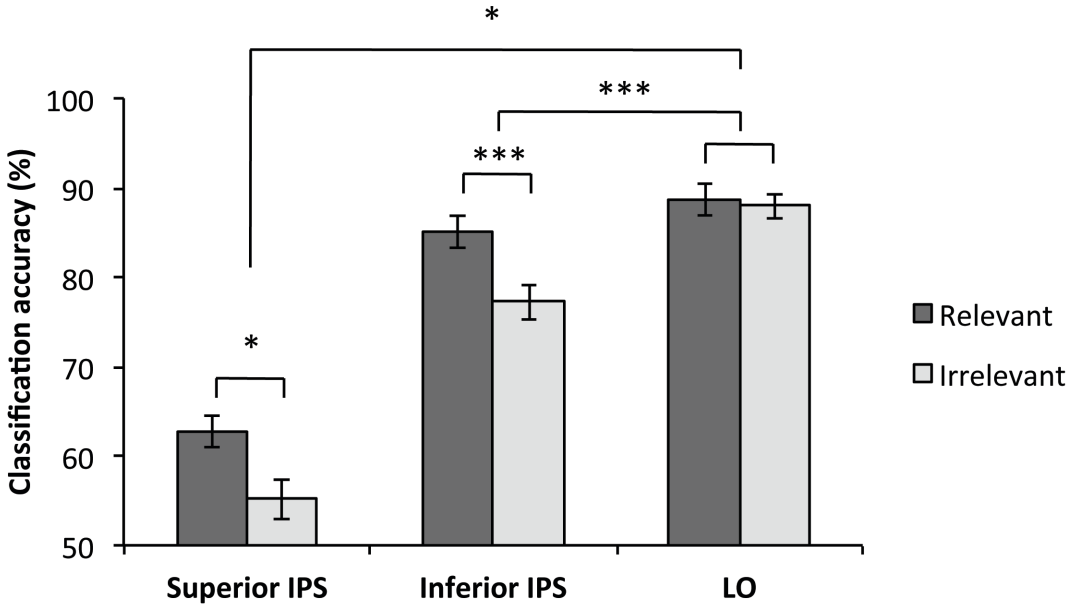

Figure 6. Decoding accuracies of task-relevant (dark grey bars) and task-irrelevant (light grey bars) feature dimensions in superior IPS, inferior IPS, and LO. Relevant features were shape in the shape and

conjunction 1-back tasks, and location in the location and conjunction 1-back tasks. Irrelevant features were shape in the location 1-back and motion detection tasks, and location in the shape 1-back and motion detection tasks. Relevant features were decoded significantly better than irrelevant features in superior and inferior IPS, but not in LO, showing task-dependent representation in these parietal regions. Error bars indicate within-subject standard error of the mean. * $p < .05$; ** $p < .01$; *** $p < .001$.

A repeated measures ANOVA with region (superior IPS, inferior IPS, and LO) and task-relevancy (relevant vs. irrelevant) as factors revealed a main effect of region ($F_{(2,26)} = 66.075$, $p < .001$), with overall lower decoding accuracy in superior IPS than in inferior IPS and LO ($t_{(13)} = 8.044$, $p < .001$, and $t_{(13)} = 12.667$, $p < .001$, respectively, Bonferroni corrected) (Figure 6). There was also a main effect of task, with task-relevant dimensions decoded better than task-irrelevant dimensions ($F_{(1,13)} = 12.285$, $p = .004$). Importantly, interaction between region and task-relevancy was significant ($F_{(2,26)} = 6.552$, $p = .005$). Further pairwise comparisons revealed that task-relevant decoding showed significantly greater accuracy than task-irrelevant decoding in superior and inferior IPS ($ts > 2.717$, $ps < .018$). In addition, this task-relevancy effect was significantly greater in superior and inferior IPS than in LO (region by task-relevancy interactions, $Fs > 7.271$, $ps < .018$).

Taken together, we found visual information representation was decoded better when it was task relevant in superior and in inferior IPS, but not in LO. This task-driven visual information representation in superior and inferior IPS is similar to responses observed in LIP neurons in monkey neurophysiology studies (Fitzgerald, Swaminathan, & Freedman, 2012; Freedman & Assad, 2006; 2009; Swaminathan & Freedman, 2012; Toth & Assad, 2002).

Shape and Location decoding

Next, we examined shape and location decoding separately in each task to determine which feature dimension contributed to task-dependent representation. First, we tested shape and location decoding performance in each task in superior IPS. In the shape 1-back task, object shape, rather than its location, became task relevant. We found significant shape decoding ($t_{(13)}$ = 3.329, p = .005) and marginally significant location decoding in the shape 1-back task ($t_{(13)}$ = 1.82, p = .092) (Figure 7a). In contrast, in the location 1-back task, location, but not shape, could be decoded ($t_{(13)}$ = 2.722, p = .017, and $t_{(13)}$ = 1.532, p = .125, respectively). In the conjunction 1-back task where both shape and location were task-relevant, we observed successful decoding of both shape and location (ts > 2.747, ps < .017). In the motion detection task, neither shape nor location was task-relevant and we failed to find any information representation ($t_{(13)}$ < 1, p = .913 for shape, $t_{(13)}$ = 1.147, p = .272 for location).

To determine whether shape and location representations were modulated by task demands, we compared task-relevant and task-irrelevant shape (or location) decoding accuracies. In superior IPS, shape decoding accuracy in relevant tasks (shape and conjunction 1-back tasks) were greater that that in irrelevant tasks (location 1-back and motion tasks), showing task-dependent shape representation ($t_{(13)}$ = 2.14, p = .052). However, decoding accuracies of location in relevant and irrelevant tasks were not different from each other (location decoding in the location and conjunction tasks vs. that in the shape and motion tasks, $t_{(13)}$ = 1.37, p = .192).

Thus, superior IPS carried shape representation, but only when such information was required by the task. Significant location decoding was found in the location and conjunction 1-back tasks, where such information was required. Nevertheless, task-dependent modulation of

location was not as dynamic as shape representation, as location decoding accuracy did not differ significantly whether tasks required location information or not.

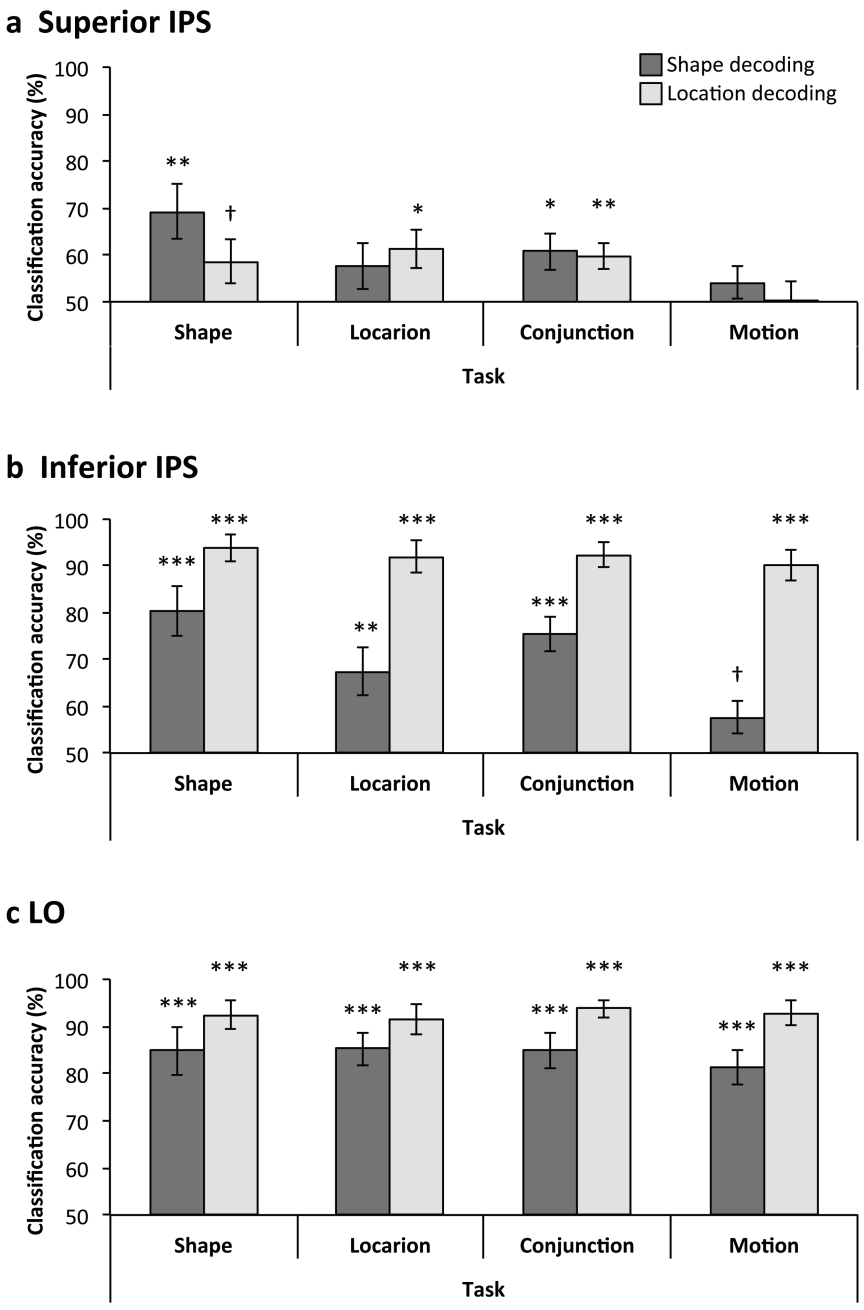

Figure 7. Decoding accuracies of shape and location in each task in (a) superior IPS, (b) inferior IPS and (c) LO. Dark and light grey bars indicate shape and location decoding, respectively. Decoding accuracies

were compared to the chance level (50%). X axis shows shape, location, conjunction, and motion tasks. Error bars indicate within-subject standard error of the mean.
(a-b) In superior and inferior IPS, shape decoding accuracies in the shape and conjunction 1-back tasks were greater than those in the location 1-back and motion tasks, showing task-dependent shape encoding. However, location decoding accuracies did not differ between tasks. **(c)** In LO, both shape location could be decoded in all four tasks, but task did not modulate shape and location representations by task. † $p < .1$; * $p < .05$; ** $p < .01$; *** $p < .001$.

Next, we evaluated shape and location decoding in inferior IPS in each task. Inferior IPS revealed strong location representation across all four tasks ($ts > 11.649$, $ps < .001$) (Figure 7b). Significant location decoding in inferior IPS is consistent with its role in object individuation suggested by previous studies (Xu, 2009; Xu & Chun, 2006). Interestingly, object shape also could be decoded in all four tasks ($ts > 3.397$, $ps < .005$ in the shape, location, and conjunction 1-back tasks; $t_{(13)} = 2.144$, $p = .051$ in the motion detection task). Though shape could be decoded even in tasks that did not require such information, the decoding accuracy was significantly lower than that in tasks required shape information (shape decoding in the shape and conjunction tasks vs. in the location and motion tasks, $t_{(13)} = 5.448$, $p < .001$). Location decoding did not show any difference whether task required location encoding or not ($t < 1$, $p = .907$). Thus, inferior IPS also showed task-dependent object shape, but not location, representation as superior IPS did.

To test whether task-dependent modulation of shape representation is prevalent in the visual system, next we examined LO, a visual object processing region. In LO, both object shape and location information could be reliably decoded in all four tasks ($ts > 16.348$, $ps < .001$, Figure 7c), consistent with previous findings (Kravitz, Kriegeskorte, & Baker, 2010; Schwarzlose, Swisher, Dang, & Kanwisher, 2008). Both shape and location decoding did not differ between relevant and irrelevant tasks, showing task-independent object representation in LO ($ts < 1$, $ps > .551$).

These results suggest that task-dependent modulation of visual representation found in superior and inferior IPS was caused by better shape decoding in tasks that required such information.

2.3 Discussion and conclusion

Taken together, we found that functional sub-regions in human IPS carry visual shape representation in a task-dependent manner. Across four tasks with identical visual input, superior IPS discarded irrelevant shape feature of the attended object when such information was not required by task, consistent with previous findings that showed selective encoding of task-relevant information in superior IPS (Jeong & Xu, 2013; Xu, 2010) and flexible information processing in human parietal cortex (Liu, Hospadaruk, Zhu, & Gardner, 2011; Thompson & Duncan, 2009; Woolgar, Hampshire, Thompson, & Duncan, 2011). Inferior IPS, an adjacent region along IPS, encoded shape representation regardless of task demands. However, shape was decoded better when it was task-relevant in inferior IPS, showing task-dependent modulation.

Interestingly, object location representation was not dynamically modulated by task demands in both superior and inferior IPS. This could be due to obligatory processing of location information during object processing. Theories on visual object perception and recognition suggest that location encoding is the initial stage of visual object processing. For example, according to feature integration theory, focal attention needs to bind location and features to enable identification of an object with multiple features (Treisman & Gelade, 1980). Similarly, object-file theory suggests that objects are first individuated based on location and then their

detailed features are identified (Kahneman et al., 1992; Xu & Chun, 2009). Also, in the current study, participants could not predict the exact position of the object as it appeared in a different position in every trial. Thus, they had to locate the object as the first step of visual information processing in all four tasks, resulting in task-independent location representation.

The task-dependent shape representation in human IPS is similar to the response properties of LIP neurons previously reported in monkey neurophysiological studies. For example, LIP neurons have exhibited selectivity for shape (Sereno & Maunsell, 1998), color (Toth & Assad, 2002), task sets (Stoet & Snyder, 2004), and category membership (Freedman & Assad, 2006; Swaminathan & Freedman, 2012). Although the exact human homologue of LIP is still under debate, previous studies on visual information encoding in superior IPS (Christophel et al., 2012; Jeong & Xu, 2013; Xu, 2010; Xu & Chun, 2006; 2009) and our current results suggest possible functional correspondence between the human superior IPS and monkey LIP.

Growing evidence suggests that human parietal cortex is involved in a network of regions that enables flexible information processing (Cole et al., 2013; Fedorenko, Duncan, & Kanwisher, 2013; Vincent et al., 2008). The current results showed human IPS may be one of the key regions in the control network that supports the moment-to-moment visual information processing in a task-dependent manner.

2.4 Methods

Participants

Fourteen participants took part in the study including the author (8 females, mean age 29.5, SD 3.52). One additional participant was scanned, but excluded from further data analysis due to a failure to maintain proper fixation. All participants were right-handed and had normal or corrected-to-normal visual acuity. They were recruited from the Harvard University community, gave an informed consent prior to participation and received payments. The experiment was approved by the Harvard University Committee on the Use of Human Subjects.

Experimental design

Main experiment

The main experiment included four tasks; shape 1-back, location 1-back, conjunction 1-back, and motion detection task. The display contained a white-colored square that subtended 10.5° x 10.5° in the center on a grey-background (see Figure 5). In each block, ten exemplars from one object category were shown sequentially in one visual field within the white-colored square. Gray-scaled photographs of shoes and bikes with a side view were used as stimuli (see Figure 5) and different object categories were viewed in different trial blocks. Each item subtended approximately 5° x 2.6°. Though all the exemplars within a block were presented in one visual field, the exact position of the item within the visual field varied slightly. Specifically, there were four possible positions in each visual field; upper left, upper right, lower left, and lower right, with the distance between adjacent positions being approximately 0.9° horizontally and vertically apart. In the shape 1-back task, participants were asked to pay attention only to identity repetition and ignore positions. In the location 1-back task, they were asked to pay attention to position repetition while ignoring identities. In the conjunction 1-back task, they

monitored both identity and position repetition. To ensure participants pay attention to both dimensions, catch trials with only identity or position repetition were included in the conjunction 1-back task. Participants had to press a button when an immediate repetition of the attended feature(s) occurred. In the motion detection task, participants monitored an occasional spatial jitter (either horizontal or vertical) of the item and pressed a button corresponding to the direction of the jitter. The repetition (or motion) occurred once in half of the stimulus block and twice in the other half of the stimulus blocks. Thus, participants had to pay attention even after they detected the first repetition/motion, because there could be second repetition/motion. The order of the four tasks was counterbalanced across participants.

Sixteen stimulus blocks were included in each run (2 object categories x 2 locations x 4 repetitions). The order of the stimulus blocks and the order of the images within each block were randomly decided in each run. Each stimulus block lasted 8 sec and contained 10 images, with each image appearing 300 ms followed by a 500 ms blank display. Fixation block, which lasted 8 sec, was inserted at the beginning and end of the run, and between each stimulus block. Each participant received 4 runs for each task, each lasting 4 min 32 sec.

To ensure proper central fixation of each participant, we monitored eye movements in the main experiment with an EyeLink 1000 eye tracker.

Inferior IPS/LO localizer

Participants viewed blocks of objects and noise images (both subtended approximately 12°x 12°). The object display contained four objects that appeared above, below, left, and right to the fixation (the distance between the fixation and the center of each object was 4°). Gray-scaled

photographs of everyday objects including those appeared in the main experiment (e.g., shoes, bikes, guitars, couches, and so on) were used as stimuli. The objects appeared on white-colored placeholders (4.5° x 3.6°) that were visible during the object image block. The noise images were phase-scrambled version of the same object images. Each block lasted 16 sec and contained 20 images, with each image appearing for 500 ms followed by a 300 ms blank display. The participants were asked to detect the direction of a slight spatial jitter (either horizontal or vertical), which occurred randomly once in every 10 images. Eight object blocks and eight noise blocks were included in each run. Each participant conducted two or three runs, each lasting 4 min and 40 sec.

Superior IPS localizer

To localize the superior IPS, we conducted an object VSTM experiment. Participants were asked to remember category and location of objects in a sample display. The sample display contained 1 to 4 objects appearing in 4 possible locations. Each item in the sample display was chosen from a different category. After a short delay, a probe item (a new object) was shown at one of the locations previously occupied by sample items. For no-change trials, the probe item matched the category of the sample item shown at the same location in the sample display. For change trials, the probe item was an exemplar from a different category. Half of trials were change trials. Gray-scaled photographs of objects from four categories (shoe, bike, guitar, and couches) were used as stimuli. The size of the individual object and the whole display were identical to those used in inferior IPS/LO localizer experiment.

Each trial began with a fixation period (1,000 ms), followed by a sample display (200 ms), a delay period (1,000 ms), and a test display/response period (2,500 ms), and a feedback (1,300 ms). With a counterbalanced trial history design, there were 15 stimulus trials for each set size as well as 15 fixation trials in which only the fixation dot appeared for 6 sec. Three filler trials were added for practice and trial history balancing, but were excluded from data analysis. Each participant was tested with two runs, each lasting 8 min.

fMRI methods

fMRI data were acquired from a Siemens Tim Trio 3T scanner at the Harvard Center for Brain Science in Cambridge, MA. Participants viewed images back projected onto a screen at the rear of the scanner bore through an angled mirror mounted on the head coil. All experiments were controlled by an Apple MacBook Pro running Matlab with Psychtoolbox extensions (Brainard, 1997). For anatomical images, High-resolution 144 T1-weighted images (echo time, 1.54 ms; flip angle, 7°; 256 x 256 matrix size; repetition time, 2,200 ms; 1 mm x 1 mm x 1 mm voxel size) were acquired. For the main experiment and inferior IPS/LO localizers, thirty-one slices of 3 mm thick (3 mm x 3 mm in plane, 0 mm skip) T2*-weighted images were acquired using standard protocols. T2*-weighted image parameters were; echo time 30 ms; flip angle, 90°; 72 x 72 matrix; repetition time, 2,000 ms (136 volumes for the main experiment, 140 volumes for inferior IPS/LO localizer runs). For superior IPS localizer runs, twenty-four slices of 5 mm thick (3 mm x 3 mm in plane, 0 mm skip) images parallel to AC-PC line were acquired (320 volumes; echo time, 29 ms; flip angle, 90°; 72 x 72 matrix; repetition time, 1,500ms).

Data analysis

fMRI data were analyzed on native space with BrainVoyager QX (http://www.brainvoyager.com). 3-D motion correction, slice acquisition time correction, and linear trend removal were conducted during data preprocessing.

LO and inferior IPS ROIs were defined as voxels showing higher activations to the objects than to the noise displays (false discovery rate (FDR) $q < .05$, corrected for serial correlation) in lateral occipital cortex and IPS, respectively. Following previous studies (Todd & Marois, 2004; Xu & Chun, 2006), superior IPS was defined as voxels that tracked each participant's behavioral VSTM capacity. VSTM capacity of each participant was estimated using Cowan's K formula (Cowan, 2001). To define the superior IPS ROI, we performed multiple regression analysis with the regression coefficients for each VSTM set size weighted by each participant's behavioral VSTM capacity for that set size. Superior IPS was defined as voxels showing significant activations in the regression analysis (FDR $q < .05$, corrected for serial correlation).

We overlaid ROIs onto the data from the main experiments and extracted each voxel's β-weights for each stimulus condition from each ROI. To decode shape and location representations in each ROI in each task, we used a linear support vector machine (SVM) (Chang & Lin, 2011). Each voxel's β-weight was normalized using z score. We normalized the data to remove possible contribution of response amplitude differences among tasks. This normalization did not have a significant influence as non-normalized data also showed similar results. We divided the four runs in each task into three training runs and one test run, with N-fold cross validation. Linear classifiers were trained to predict shape (shoe vs. bike) and location (up vs.

down) in each task in each ROI separately. Decoding accuracy was compared to the chance level (50%) using one sample t-test.

To evaluate the existence of task-dependent representation across ROIs, we averaged decoding accuracies of task-relevant and -irrelevant feature dimensions separately. Task-relevant dimensions were shape in the shape and the conjunction 1-back tasks, and location in the location and the conjunction 1-back tasks. Task-irrelevant dimensions were shape in the location 1-back and the motion detection tasks, location in the shape 1-back and the motion detection tasks. Repeated measures ANOVA with region (superior IPS, inferior IPS, and LO) and task-relevancy (relevant vs. irrelevant dimensions) as factors, and paired samples t-test were performed to evaluate task-dependent representation within and between ROIs.

All statistical tests were two-tailed with a significance level of .05.

3

Abstract object identity representation in human superior intra-parietal sulcus

3.0 Abstract

Although the primate parietal cortex has been traditionally associated with spatial location and attention-related processing (Colby & Goldberg, 1999; Gottlieb & Balan, 2010), using fMRI and multi-voxel pattern analysis (MVPA), here we show that highly abstract object identity information can be robustly represented in human superior intra-parietal sulcus (IPS), a parietal region previously shown to track the content of visual short-term memory (VSTM) (Todd & Marois, 2004; Xu & Chun, 2006). We presented human observers with face images from well-known movie actors and asked them to extract face identities from these images. Despite large

variations in viewpoint, hairstyle, facial expression and age, in superior IPS, we found higher correlations of fMRI response patterns between two sets of face images belonging to the same than different actors, indicating abstract face identity representation in this brain region. Such identity representation was not limited to faces and could also be seen for well-known cars when they were shown embedded in different scenes, viewpoints and sizes. Critically, identity representation in superior IPS closely tracked perceived object identity similarity in behavioral measures, supporting its role in goal-directed visual information processing. Meanwhile, none of the ventral and lateral visual object processing regions we examined exhibited such representation. The human parietal cortex thus plays a greater role than simply directing attentional resources during visual perception as is commonly known. But rather, a sub-region within parietal cortex can directly represent incoming visual information as abstract as object identity. We propose that human superior IPS functions similarly as the random access memory (RAM) in a computer and can flexibly represent a variety of task-relevant visual information to support goal-directed visual information processing in the brain.

3.1 Introduction

Decades of cognitive neuroscience research has attributed the function of the primate parietal cortex primarily to spatial location and attention-related processing (Colby & Goldberg, 1999; Gottlieb & Balan, 2010). Emerging evidence from monkey single neuron recording studies, however, shows that object information, such as color, shape, motion, and category membership,

can be directly represented in lateral intra-parietal (LIP) sulcus (Fitzgerald et al., 2011; Freedman & Assad, 2006; Sereno & Maunsell, 1998; Toth & Assad, 2002). Consistent with these neurophysiological findings, human fMRI studies have reported the encoding of basic visual features, such as color and shape, along IPS (Christophel et al., 2012; Konen & Kastner, 2008; Todd & Marois, 2004; Xu & Chun, 2006). Superior IPS, in particular, has been linked to the encoding and storage of object color and shape in VSTM in a task dependent manner (Jeong & Xu, 2013; Todd & Marois, 2004; Xu, 2010; Xu & Chun, 2006; 2009; Xu & Jeong, in press). This collection of evidence calls for a revision of our understanding of the role of the primate parietal cortex in visual processing and suggests that this brain region plays a critical role in task-driven visual representation.

In everyday vision, task-relevant visual information varies drastically across tasks, ranging from simple features, such as color and shape, to high-level ones, such as abstract object identity representation invariant to changes in view point, size and other non-essential visual features. Although the representation of abstract object identity is fundamental to human vision, whether or not it can be directly represented in the primate parietal cortex has not been shown. If the parietal cortex plays a critical role in task-driven visual representation, then it must be capable of representing a great variety of visual information, including abstract object identities. Moreover, such abstract object identity representation must be directly linked to behavior. In the studies presented here, we provide evidence supporting both of these predictions.

Among the many object identities we extract in everyday vision, face identity is perhaps the most challenging one to form, owning to the greater amount of changes that could be associated with faces without changing their identities. This includes changes such as viewpoint,

expression, hairstyle, and age. Reflecting this computational challenge, a specialized brain network has been dedicated for face processing (Tsao, Moeller, & Freiwald, 2008). Within this network, face-identity representation has been reported in regions surrounding the right fusiform face area (FFA) and in anterior inferior temporal cortex in both neuropsychological and fMRI MVPA studies (Anzellotti, Fairhall, & Caramazza, 2013; De Renzi, Perani, Carlesimo, Silveri, & Fazio, 1994; Kriegeskorte, Formisano, Sorger, & Goebel, 2007; Nestor, Plaut, & Behrmann, 2011). However, whereas real-world faces vary freely across multiple dimensions, these fMRI studies have only employed limited manipulations, changing only viewpoint or expression, and have not revealed the representation of real-world abstract face identity in these ventral brain regions. Regardless, given the importance of face identity representation in everyday vision and social interactions, parietal cortex ought to carry these representations robustly if it were to play a critical role in task-driven visual processing. Thus, to provide the most stringent test on parietal cortex's ability to represent abstract object identities, we tested its ability to represent face identity from real-world face images varying freely in viewpoint, expression, hairstyle, and age. We used fMRI MVPA, a tool that has been widely used in recent fMRI studies to understand information representation in the brain (Haxby et al., 2001; Norman et al., 2006).

3.2 Experiment 1 results

In Experiment 1, we used face images of Leonardo DiCaprio and Matt Damon, two well-known actors matched in overall appearance. To encourage the formation of real-world abstract

face identity representations, we varied viewpoint, hairstyle, facial expression, and age of the faces and constructed two unique face sets for each actor. While lying in an MRI scanner, observers viewed the sequential presentation of the images in each face set multiple times and detected an occasional presence of an oddball face drawn from one of eight other male actors (Figure 8a and Figure 12). The formation of abstract face identity for the two actors was thus necessary to ensure successful task performance. To remove the contributions of oddball face responses to the observed fMRI response patterns, they were excluded from data analysis (see Methods). In addition to face images, we also showed the written names of each actor in different fonts in the same oddball task (Figure 8b). The name task allowed us to evaluate whether brain responses associated with face images reflected a visual code (i.e., abstract face identity representation) or a phonological code (i.e., observers rehearsing an actor's name), as phonological code is automatically activated during word reading (Van Orden, 1991). Although names can evoke face identity representations, because our oddball name task could be performed based on the phonological code alone, the activation of face identity representations would be unnecessary here.

We obtained averaged fMRI response patterns for each face set (or name set) of each actor in independently defined brain regions of interest (ROIs, see Figure 8c and Methods). We targeted our investigation in the parietal cortex to superior IPS, as this brain region tracks the encoding and storage of simple visual features in VSTM in a task-dependent manner (Todd & Marois, 2004; Xu & Chun, 2006), and is thus a promising parietal region where abstract face identity representation may exist. Besides superior IPS, we also examined representation in three ventral brain regions, one involved in object shape representation - the lateral occipital region

(LO) (Malach et al., 1995), one involved in face processing - the right FFA(Kanwisher, McDermott, & Chun, 1997), and the third involved in letter string processing - the visual word form area (VWFA) (L. Cohen et al., 2000).

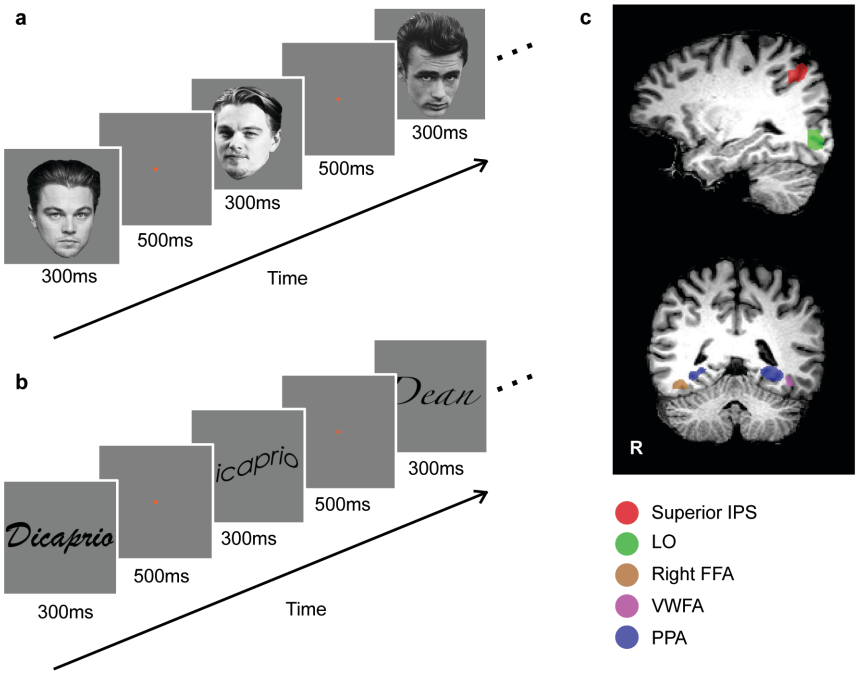

Figure 8. Experiment 1 example stimuli, trial structure and example ROIs. a, Example images from a block of face trials. Face images from two well-known actors, Leonardo DiCaprio and Matt Damon, were used here. Within a block of trials, observers viewed a sequential presentation of ten face images sharing the same identity but differed in viewpoint, hairstyle, facial expression, and age. Observers were asked to detect the occasional presence of an oddball face from one of eight other actors. James Dean's face is shown here as the oddball among Leonardo DiCaprio's faces. **b,** Example images from a block of name trials. Actor names written in different fonts were used here in an oddball name detection task. James Dean's name is shown here as the oddball among Leonardo DiCaprio's names. Oddball occurred rarely and blocks containing the oddball were removed, leaving only blocks containing face or name images of the same actor included in the final analysis. **c,** Example ROIs from one representative observer.

Similar to the approach used by Haxby and colleagues (2001), we correlated fMRI response patterns obtained from different face sets in each ROI and Fisher-transformed the resulting correlation coefficient (see Methods). In superior IPS, two face sets from the same actor elicited a higher correlation than two each from a different actor (Figure 9b, paired samples t-test,

two-tailed, $t_{(12)} = 2.86$, $P = .014$; this applies to all subsequent analyses except where noted). Thus, despite large variations in face appearance, two distinctive face sets sharing an identity were represented more similarly than two that differed in identity, indicating the representation of abstract face identities in superior IPS. Such representation, however, was not found in LO, the right FFA, or VWFA (*ts* < 1.13, *P*s > .27; see Figure 9b). Further pairwise comparisons revealed that superior IPS differed significantly from the other ROIs in abstract face identity representation (region by identity interaction, *F*s > 8.71, *P*s < .012). Differences among the brain regions could not be attributed to voxel number differences, as both superior IPS and LO contained similar number of voxels (see Table 1). Additionally, when the number of voxels in each ROI was limited to 50, the same results were obtained (see Figure 13a).

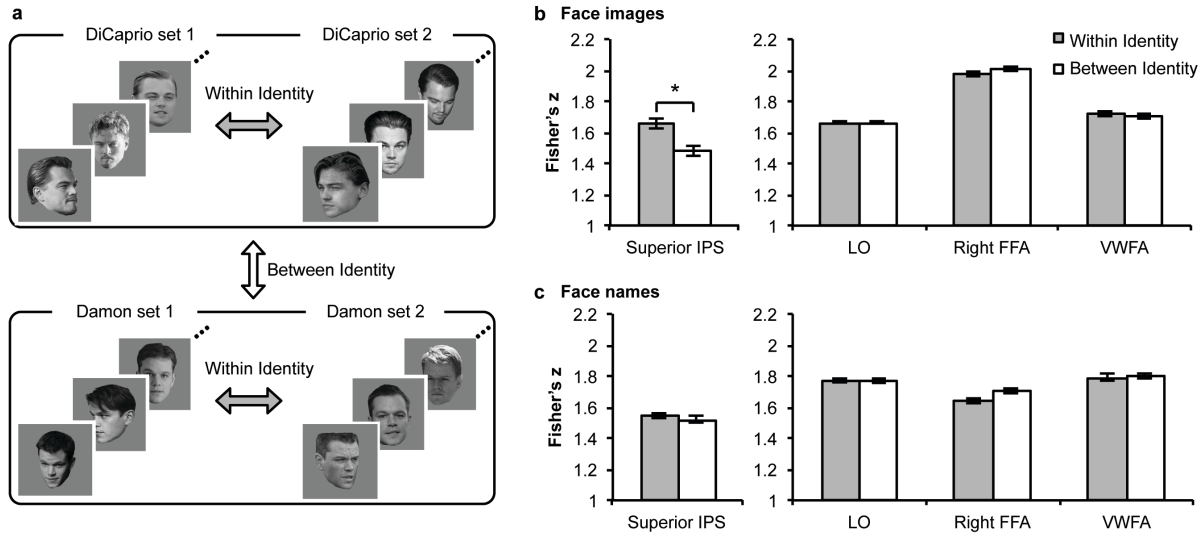

Figure 9. Experiment 1 design and results. a, Schematic illustration of the key comparisons made in the experiment. To evaluate the existence of abstract identity representation, we examined whether within-identity correlation was greater than between-identity correlation. Within-identity correlation referred to the correlation of fMRI voxel response patterns between two face sets (or two name sets) from the same actor, whereas between-identity correlation referred to pattern correlation between two face sets (or two name sets) each from a different actor. **b,** Fisher-transformed correlation coefficients (z) from face sets in superior IPS, LO, the right FFA, and VWFA. Only superior IPS showed a higher within-identity than

between-identity correlation, indicating abstract face identity representation in this brain region despite large variations in viewpoint, hairstyle, facial expression and age of the face images used. **c,** Fisher-transformed correlation coefficients from name sets in the same brain regions. None of the regions showed a higher within-identity than between-identity correlation, indicating the absence of identity representation in these brain regions when name stimuli were used. Grey bars indicate within-identity correlations and white bars indicate between-identity correlations. Error bars indicate within-subject standard error of the mean. * $P < .05$.

Table 1. Number of voxels in each ROI (mean and SD).

	Superior IPS	LO	Right FFA	PPA	VWFA
Experiment 1 (faces, 2 actors)	260.61 (115.19)	303.38 (51.12)	70.46 (28.23)	-	68.3 (29.25)
Experiment 2 (cars)	271.61 (70.13)	275.07 (46.60)	-	227.23 (63.51)	64.15 (27.88)
Experiment 3 (faces, 8 actors)	276.09 (81.86)	322.8 (59.45)	69 (23.87)	-	-

Face identity representation in the superior IPS could not be attributed to perceptual differences among the face sets, as low-level perceptual differences such as luminance and spatial frequency could not account for our results (see Appendix A). Moreover, although some of the ventral ROIs showed sensitivity to perceptual differences among the sets (i.e., showing a higher correlation between odd and even runs of the same set than between different sets sharing an identity, see Appendix A and Figure 14a), none of them showed the same face identity effect as superior IPS.

Face identity representation in superior IPS reflects the representation of an abstract visual code and not that of a phonological code generated by observers actively rehearsing the actors' name while viewing the face images. This is because in our name task no name identity representation was found (i.e., no difference in correlation between two sets that shared name identity and those that did not, $t_{(12)} < 1$, $P = .56$ in superior IPS, ts < 1, Ps > .84 in LO and VWFA, and $t_{(12)} = -2.06$, $P = .066$ in the right FFA in the opposite direction; see Figure 9c). Further

comparison between tasks revealed that in superior IPS identity representation was marginally greater for faces than for names (task by identity interaction, $F_{(1, 12)} = 4.06$, $P = .067$).

Thus, among the brain regions examined, superior IPS was the only one that showed sensitivity to face identity change despite large variations in face appearance within each fact set. To our knowledge, this is the first evidence showing the existence of real-world abstract face identity representation in the human parietal cortex.

3.4 Experiment 2 results

Abstract object identity representation is not limited to the perception of faces but applies to the perception of visual objects in general. To replicate and generalize our findings, in Experiment 2, photographs of two familiar car models, BMW Mini and Volkswagen Beetle, were used. Images from these car models were shown in different viewpoints, sizes, and background scenes as how they would naturally appear in everyday visual perception (Figure 12b). As in our face experiment, the written names of the cars were also shown in different fonts. Using the same oddball detection task, abstract car identity representation was examined in superior IPS, LO, and VWFA. As cars were shown embedded in background scenes, responses were also examined in the parahippocampal place area (PPA), a brain region specialized in scene processing (Epstein & Kanwisher, 1998).

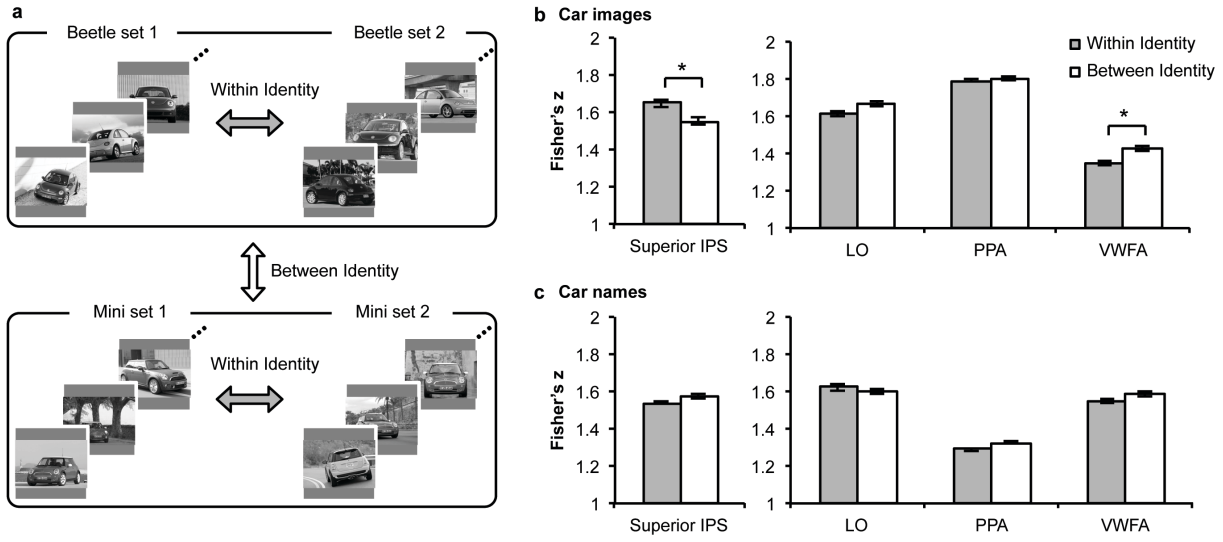

Figure 10. Experiment 2 design and results. a, Schematic illustration of the key comparisons made in the experiment. Similar to faces and names in Experiment 1, we examined for car images and names whether within-identity correlation was greater than between-identity correlation. **b,** Fisher-transformed correlation coefficients (z) for car image sets in superior IPS, LO, PPA, and VWFA. Replicating the results for faces in Experiment 1, only superior IPS showed a higher within-identity than between-identity correlation, indicating abstract car identity representation in this brain region despite large variations in viewpoint, size, and the background scene in which the car appeared. **c,** Fisher-transformed correlation coefficients for car name sets in the same brain regions. As with face names in Experiment 1, none of the regions showed a higher within-identity than between-identity correlation, indicating the absence of identity representation in these brain regions when name stimuli were used. Grey bars indicate within-identity correlation and white bars indicate between-identity correlation. Error bars indicate within subject standard error of the mean. * $P < .05$.

Replicating the results of the face experiment, only superior IPS revealed an abstract car identity representation (Figure 10b), showing higher correlation in fMRI response patterns between two sets of car images sharing the same than different identities ($t_{(12)} = 2.26$, $P = .043$). Such real-world abstract car identity representation, however, was not found in the other ROIs examined ($t_{(12)} < 1$, $P > .62$ in PPA; $t_{(12)} = -1.96$, $P = .073$ in LO, $t_{(12)} = -2.8$, $P = .016$ in VWFA, both in the opposite direction). Further pairwise comparisons revealed that superior IPS differed significantly from the other ROIs in car identity representation (region by identity interactions, Fs > 7.87, Ps $< .016$). As in the face experiment, differences among the brain regions could not be

attributed to voxel number differences, as voxel numbers were similar in superior IPS and LO (Table 1), and the same results were obtained when the number of voxels in each ROI to was limited to 50 (Figure 13c). Car identity representation in superior IPS also could not be accounted for by perceptual differences among the sets (see Appendix A and Figure 14c).

Meanwhile, no car name identity representation was found in any of the ROIs examined (Figure 10c, *t*s < 1, *P*s > .13 in superior IPS, LO, and VWFA; $t_{(12)}$ = -1.96, *P* = .073 in PPA in the opposite direction). Comparison between tasks revealed that identity representation was greater for car images than for car names in superior IPS (task by identity interaction, $F_{(1,12)}$ = 9.23, *P* = .01). These results thus replicated those of the face experiment and showed that abstract identity representation exists in superior IPS for both faces and non-face objects such as cars.

3.5 Experiment 3 results

In Experiment 1, the decoding of face identity was examined between two individuals. To generalize our finding beyond these two specific individuals, in Experiment 3, the same oddball detection task paradigm was applied to face images from eight famous actors (see Figure 11a, Figure 12c, and Methods). Replicating the results from Experiment 1, when all pairwise comparisons between the 8 actors were averaged, face identity decoding was again observed in superior IPS ($t_{(10)}$ = 2.58, *P* = .027, Figure 11b), but not in LO or the right FFA (*ts* <1.17, *Ps* > .266), with greater identity decoding in superior IPS than LO (region by identity interaction, $F_{(1,10)}$ = 8.13, *P* = .017). When the number of voxels in the two regions were matched (up to 50 most responsive voxels), the difference between superior IPS and the right FFA approached

significance ($F_{(1,10)} = 4.07$, $P = .071$, see Figure 13e). Such face identity decoding in superior IPS was not driven by the decoding of the best face pairs, as removing the two best face pairs from the analysis did not change the results ($t_{(10)} = 2.231$, $P = .049$). Similarly, removing the two worst pairs from the analysis did not improve the results in LO and the right FFA with decoding still be at chance ($ts < 1.324$, $Ps > .215$).

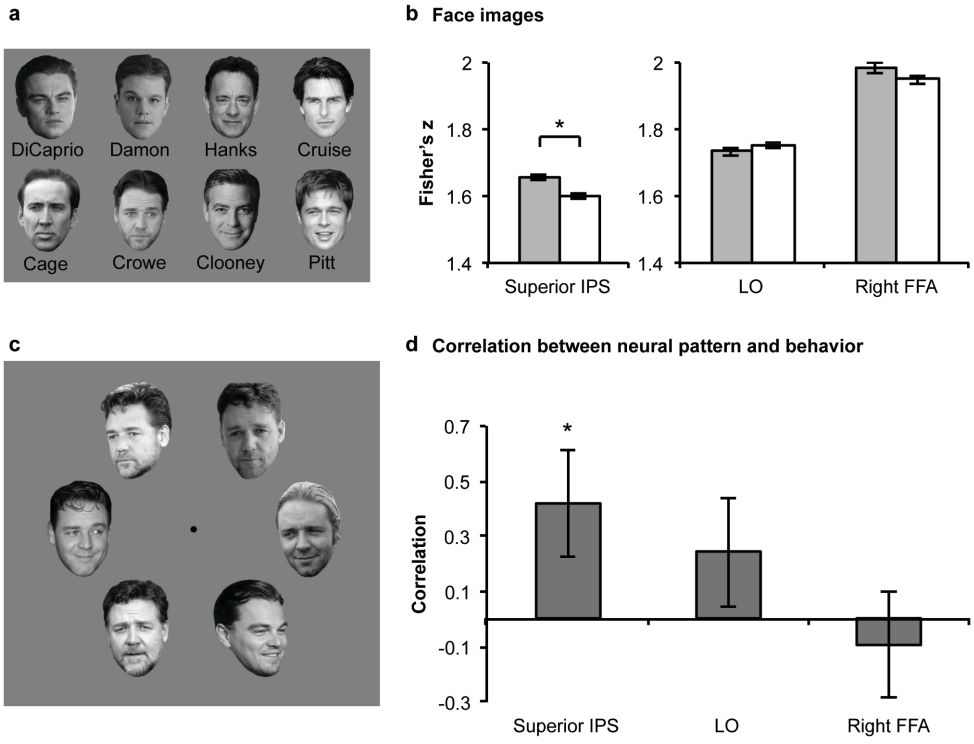

Figure 11. Experiment 3 example stimuli and results. a, Example images of the eight actors used. **b,** Fisher-transformed correlation coefficients (z) between face image sets in superior IPS, LO, and the right FFA. Superior IPS again showed higher within- than between-identity correlation whereas LO and right FFA did not. Grey bars indicate within-identity correlation and white bars indicate between-identity correlation. **c,** An example face visual search display. Observers performed a speeded search for the presence of the face of a target actor among faces of a distractor actor. Target face appeared in 50% of the trials. In the example shown, DiCaprio is the target actor and Crowe is the distractor actor. **d,** Correlation between behavioral similarity measure of face identity (as measured by visual search speed) and neural similarity measure of face identity (as measures by fMRI pattern correlations) in each ROI. This correlation reached significance only in superior IPS, indicating that face representation formed there closely tracked perception. Error bars indicate within subject standard error of the mean. * $P < .05$.

The oddball detection task recruited multiple brain regions, other than superior IPS, that might contribute to identity representation. We chose ROIs based on their known functional role in visual information processing, but we could have overlooked identity representation that might exist in other parts of ventral/temporal and posterior parietal cortices. Also, lateral prefrontal cortex (LPFC) is a promising candidate region to find identity representation as this region is recruited during working memory (Goldman-Rakic, 1995) and categorization tasks (Freedman, Riesenhuber, Poggio, & Miller, 2001). To assess these possibilities, we examined face identity decoding in LPFC, PPC, and ventral/temporal cortices in Experiment 3. In this further analysis, all of these regions failed to show identity representation, confirming that such representation uniquely exists in superior IPS (see Appendix A).

Besides identity, faces may differ in other abstract properties, such as familiarity, attractiveness, trustworthiness, and so on. To examine whether decoding in superior IPS reflected face identity representation and not any of the other abstract properties associated with face perception, we compared neural measures of face similarity in superior IPS with behavioral measures of face similarity from a speeded visual search task. In addition to the fMRI study, the same observers from Experiment 3 also performed a speeded visual search task outside the scanner and searched for a target actor face among distractor faces of another actor with all the possible pairings among the 8 actors (Figure 11c). As target-distractor similarity has been shown to govern visual search efficiency (Duncan & Humphreys, 1989), search speed was used as a behavioral measure of face similarity between two actors with a slower search speed indicative of greater similarity between two face identities. From all possible pairing of the 8 face identities, we constructed a behavioral face similarity matrix. Using the fMRI correlation coefficient values, a

neural face similarity matrix was constructed separately for superior IPS, LO, and the right FFA and correlations between the behavioral and the neural measures of face similarity were then calculated (Kriegeskorte et al., 2008). This analysis revealed that behavioral similarity measure was correlated significantly with neural similarity in superior IPS ($P < .013$, permutation test), but not in LO or the right FFA ($P = .112$ and $P = .317$, respectively, see Figure 11d). Moreover, behavioral similarity measure correlated more with neural similarity measure in the superior IPS than in the right FFA ($P =. 028$, permutation test; this correlation did not differ between the superior IPS and LO, $P = .254$,). The results remained the same when only up to 50 top voxels were included in each ROI and when search data were truncated to remove outliers greater than 3 SD ($P = .028$ in superior IPS; $Ps > .293$ in LO and the right FFA, see Figure 15). Thus, perceived face identity similarity in our speeded visual search task was truthfully reflected in the neural response patterns in superior IPS, and more so than that in the right FFA. These results thus provide the strongest support showing that goal-driven face identity information can be directly represented in superior IPS.

3.6 Discussion and conclusion

Taken together, the results from three experiments demonstrated that real-world abstract object identity information invariant to large changes in object appearance could be robustly represented in human superior IPS. Although face identity representation has been reported in the fusiform gyrus previously (Anzellotti et al., 2013), no such representations were found in the ventral ROIs examined here, including the right FFA. This could be due to the greater perceptual

diversity of the face images used here than what was used previously, although such diversity is quite common in real world object perception. It is likely that real-world abstract object identity representation exists in more anterior ventral regions, such as anterior temporal cortex (Anzellotti et al., 2013; Nestor et al., 2011), but they were outside of our coverage in the present study.

Our results are in line with previous neurophysiological and neuroimaging findings showing that the primate parietal cortex participates in task-relevant visual information processing (Fitzgerald et al., 2011; Jeong & Xu, 2013; Liu et al., 2011; Toth & Assad, 2002; Woolgar et al., 2011; Xu, 2010; Xu & Jeong, in press). While previous studies have only documented the representation of basic visual features in the human parietal cortex, here we show for the first time that visual information as abstract as real-world face and car identity can be robustly represented in superior IPS despite large variations in face and car appearance. Importantly, we also show that this representation closely tracked the perceived object identity.

Although the frontal and parietal regions have long been implicated in task set and cognitive control (Dosenbach et al., 2008; Duncan, 2010; Duncan & Owen, 2000; Miller & Cohen, 2001), with the fronto-parietal brain network capable of rapidly updating their pattern of global functional connectivity according to task demands (Cole, et al., 2013), the manner in which task-driven visual information is processed remains poorly understood. Here we show that human superior IPS functions similarly as the RAM in a computer and represents a variety of object identities following task demands. This capability likely enables superior IPS to play a vital role in the fronto-parietal brain network in extracting task-relevant visual information to support the moment-to-moment goal-directed information processing in the brain.

3.7 Methods

Participants

Thirteen (9 females, mean age 28.6 ± 4.7), 13 (8 females, mean age 28.2 ± 4.9), and 11 (8 females, mean age 28.64 ± 3.5) observers participated in Experiments 1 to 3, respectively. Of these observers, 3 females participated in all three Experiments, 5 (2 females) participated in both Experiments 1 and 2, 2 females participated in both Experiments 1 and 3, and 2 (1 female) participated in both Experiments 2 and 3. Besides these observers, 3, 2, and 2 additional observers were tested in Experiments 1 to 3, respectively, but were excluded from data analysis due to either excessive head motion during the experiment, a failure to localize all regions of interest, or observer's failure to keep awake during the experiment. All observers were right-handed and had normal or corrected-to-normal visual acuity. They were recruited from the Harvard University community, gave informed consent prior to participation and received payments. The experiments were approved by the Harvard University Committee on the Use of Human Subjects.

Experimental design

Main fMRI experiments

In Experiment 1, face images of two well-known actors, Leonardo DiCaprio and Matt Damon, were used as stimuli. We constructed two unique face sets for each actor, with each containing five frontal and five profile/three-quarter/profile view faces of the actor. Besides faces, we also presented each actor's last name written in 20 unique fonts and constructed two name

sets for each actor, each containing 10 unique fonts. Face and name images subtended approximately 11.5° x 8.5° and 10.5° x 3.0°, respectively. See Figure 12a for the full set of stimuli used in the experiment.

The 10 images from a given set of faces or names were presented sequentially in a 8-sec block, with each image appearing for 300 ms and followed by a 500 ms blank display. The presentation of the face and name blocks was randomly intermixed within each run. Besides face and name blocks, 8-sec long fixation blocks were inserted between successive stimulus blocks and at the beginning and end of each run. Observers viewed the face or name images and detected the presence of an oddball face or name drawn from one of eight other actors (James Dean, Daniel Day-Lewis, Robert DeNiro, Gerard Depardieu, Johnny Depp, Matt Dillon, Michael Douglas, and Robert Downey Jr.). Two face images and two name images were used for each of the oddball actors. Note that because the last names of the two target actors all started with the letter "D" (i.e., DiCaprio and Damon), oddball actors' last names all started with the letter "D" in an effort to discourage observers from attending only to the first letter of each last name instead of the entire last name in the oddball-name task.

Each run contained four face blocks and four name blocks with no oddballs and one or two face or name blocks each containing a single oddball. Blocks containing an oddball were excluded from further data analysis to remove the contribution of oddball detection. When only one oddball block was present in a run, a dummy block containing no oddball was added to ensure that all runs had the same length whether or not it contained one or two oddball blocks. The dummy block was randomly chosen from one of the face or name blocks and was removed from further analysis. Each observer was tested with 10 runs, each lasting 2 min 45 sec.

In Experiment 2, the same oddball detection paradigm was used with images and names of two distinctive car models, BMW Mini and Volkswagen Beetle. In the car images, cars were shown in different viewpoints, sizes, and background scenes as how they would naturally appear in everyday visual perception. Car images and car names subtended approximately 11.5° x 7.5° and 8.0° x 4.0°, respectively. Oddball stimuli were drawn from one of sixteen other car models (Honda Accord, Nissan Altima, Toyota Camry, Honda Civic, Toyota Corolla, Chevrolet Cruze, Nissan Cube, Volkswagen Golf, Chevrolet Malibu, Ford Mustang, Honda Odyssey, Nissan Pathfinder, Toyota Prius, Land Rover Range Rover, Mercedes-Benz Roadster, Hyundai Sonata). One car image and one name image were used for each of the oddball cars.

In Experiment 3, face images of eight famous actors (Leonardo DiCaprio, Matt Damon, Brad Pitt, George Clooney, Tom Cruise, Tom Hanks, Nicolas Cage, and Russell Crowe) were used as stimuli. These actors were the top actors rated in our behavioral familiarity ratings. The face images of DiCaprio and Damon included some used in Experiment 1 and some new ones, as no profile-view images were used in this experiment to ensure that face images from all actors were easily recognizable (see Figure 12c for the complete face images in Experiment 3). As in Experiment 1, two sets of unique face images were constructed for each actor, with five frontal and five three-quarter view faces in each face set. Oddball stimuli were frontal and three-quarter view face images from sixteen other famous actors (Christian Bale, Daniel Craig, Jude Law, Michael Douglas, Jack Nicholson, Colin Firth, Robert De Niro, Bruce Willis, Orlando Bloom, Richard Gere, Mel Gibson, Ashton Kutcher, Ben Stiller, Joseph Gordon-Levitt, Benedict Cumberbatch, and Robert Downey Jr.). One frontal-view and one three-quarter view images were used for each of the oddball actors. Name blocks were not included in this experiment. Each

run contained 16 stimulus blocks, two for each of the target actors, 1 or 2 oddball blocks, and one dummy block when there was only one oddball block present. Each observer was test with 16 runs, each lasting 5 min 4 sec. Other aspects of the design were identical to that of Experiment 1.

Superior IPS localizer

Following Todd and Marois (2004), in an event-related object VSTM experiment, observers viewed in the sample display a brief presentation of 1 to 4 everyday objects, and, after a short delay, judged whether the probe object (a new object) shown in the test display matched the category of the object appeared in the same location in the sample display. A match occurred in 50% of the trials. Gray-scaled photographs of objects from four categories were used and they were shoes, bikes, guitars, and couches. Objects could appear above, below, to the left, or to the right of the central fixation. Object locations were marked by white rectangular placeholders that were always present during the trial. The placeholders subtended 4.5° x 3.6° and were 4.0° away from the fixation (center to center). The entire display subtended 12.5° x 11.8°. Each trial contained the following: fixation (1,000 ms), sample display (200 ms), delay (1,000 ms), test display/response (2,500 ms), and feedback (1,300 ms). With a counterbalanced trial history design (Todd & Marois, 2004; Xu & Chun, 2006), each run contained 15 trials for each set size and 15 fixation trials in which only the fixation dot was present for 6 sec. Two filler trials were added at the beginning and one at the end of each run for practice and trial history balancing purposes. They were excluded from data analysis. Each observer was tested with two runs, each lasting 8 min.

LO/FFA/PPA localizer

Following Kourtzi and Kanwisher (2001), Kanwisher et al. (1997), and Epstein and Kanwisher (1998), observers viewed blocks of sequentially presented face, scene, object and scrambled object images (all subtended approximately 12.0°x 12.0°). The images used were photographs of gray-scaled male and female faces, common objects (e.g., cars, tools, and chairs), indoor and outdoor scenes, and phase-scrambled versions of the common objects. Observers monitored a slight spatial jitter which occurred randomly once in every 10 images. Each run contained four blocks of each of scenes, faces, objects, and phase-scrambled objects. Each block lasted 16 sec and contained 20 unique images, with each appearing for 750 ms and followed by a 50 ms blank display. Eight-sec long fixation blocks were included at the beginning, middle, and end of each run. Each observer was tested with two runs, each lasting 4 min and 40 sec.

Behavioral visual search experiment

In the visual search experiment, observers searched for a target actor face embedded among the faces of a distractor actor. Each observer was tested with 8 blocks of trials, with each of the 8 actors in Experiment 3 serving as the target actor for one block and the remaining 7 actors serving as the distractor actors for that block. Each actor face could appear in one of six images, with three in frontal and three in three-quarter views. Each block began with an instruction indicating the target actor for that block. Observers then viewed six faces appearing in a circular array around the fixation (see Figure 11c) and made a speeded target present/absent judgment. The target actor face appeared in 50% of the trials and were shown equally often in each of the six possible locations. For a target-present trial, the target actor face was randomly

chosen from one of the 6 face images of that actor. In each trial, the distractor actor was randomly chosen from one of the remaining seven actors with either all 6 images of that actor shown for target-absent trials or a random 5 of the 6 images shown for target-present trials. In a given block, because observers were only told the target actor identity but not a specific face image of that actor to search for (three frontal and three three-quarter view faces of target actor were randomly chosen from the image sets as target images for each block), they had to form an abstract identity representation for that actor while performing the search, similar to what they had to do during the oddball detection task in the fMRI part of the experiment.

Each block of trials contained 28 practice and 84 experimental trials (7 distractor actors x 6 locations x 2 target appearance). When observers made an incorrect response, a red unhappy face flickered at fixation for 5 sec. Incorrectly responded trials were repeated at the end of each block until correct responses were obtained for all the trials in that block. Thus response accuracy was 100% correct and only search speed was analyzed and compared with neural responses.

fMRI methods

fMRI data were acquired from a Siemens Tim Trio 3T scanner at the Harvard Center for Brain Science in Cambridge, MA. Observers viewed images back projected onto a screen at the rear of the scanner bore through an angled mirror mounted on the head coil. All experiments were controlled by an Apple MacBook Pro laptop running Matlab with Psychtoolbox extensions (Brainard, 1997). For anatomical images, high-resolution T1-weighted images were acquired (repetition time, 2,200 ms; echo time, 1.54 ms; flip angle, 7°; 144 slices; matrix size, 256 x 256; and voxel size, 1 x 1 x 1 mm). For the functional images in the main experiments and in the

LO/FFA/PPA localizers, gradient-echo echoplanar T2*-weighted images were acquired (repetition time, 2,000 ms; time to echo, 30 ms; flip angle, 90°; 31 slices; matrix size 72 x 72; voxel size, 3 x 3 x3 mm; 88 volumes for Experiments 1 and 2, 152 volumes for Experiment 3, and 140 volumes for the LO/FFA/PPA localizer). For the functional images in the superior IPS localizer, gradient-echo echoplanar T2*-weighted images with slightly different parameters were acquired (repetition time, 1,500 ms; time to echo, 29 ms; flip angle, 90°; 24 slices; matrix size 72 x 72; voxel size, 3 x 3 x 5 mm; 320 volumes). All functional slices were oriented near horizontal to optimally cover parietal, occipital, and ventral cortices. This resulted in the partial exclusion of anterior temporal and orbitofrontal cortices.

Data analysis

fMRI data were analyzed in native space with BrainVoyager QX (http://www.brainvoyager.com). Data preprocessing included 3D motion correction, slice acquisition time correction, and linear trend removal. No spatial smoothing or other data preprocessing was applied.

ROI definitions

fMRI data from the localizer runs were analyzed using general linear models. Following Todd and Marois (2004) and Xu and Chun (2006), superior IPS was defined as the collection of voxels that tracked each observer's behavioral VSTM capacity. To localize these voxels, we first obtained each observer's behavioral VSTM capacity using Cowan's K formula (2001). We then performed multiple regression analysis on the fMRI VSTM data with the regression coefficient

for each set size weighted by that observer's behavioral VSTM capacity for that set size. Superior IPS was defined as voxels in bilateral parietal cortex showing significant activations in the regression analysis (false discovery rate (FDR) $q < .05$, corrected for serial correlation). More details of this analysis can be found in Todd and Marois (2004) and Xu and Chun (2006).

Following Grill-Spector et al. (1998) and Kourtzi and Kanwisher (2001), LO was defined as the collection of continuous voxels in bilateral lateral occipital cortex showing higher activations to objects than to noise (FDR $q < .05$, corrected for serial correlation). Following Kanwisher et al. (1997), the right FFA was defined as voxels in right fusiform gyrus showing higher activation for faces than for scenes and objects (FDR $q < .05$, corrected for serial correlation). Following Epstein and Kanwisher (1998), PPA was defined as voxels in bilateral collateral sulcus and parahippocampal gyrus showing higher activations for scenes than for faces and objects.

Following Cohen et al. (2002), VWFA was localized using data from the oddball detection task and defined as voxels in the left middle fusiform gyrus showing higher activations for face names than face images in Experiment 1 or higher activations for car names than car images in Experiment 2 (FDR $q < .05$, corrected for serial correlation).

MVPA

MVPA was performed with custom-made Matlab code. We overlaid the ROIs onto the data from the main experiments, applied GLM and extracted the beta-weight for each stimulus set in each voxel of each ROI. To measure identity representation in each ROI, we compared correlation coefficient between voxel response patterns from stimulus sets that shared the same

identity (within-identity correlation) with that from stimulus sets that did not share identity (between-identity correlation) (see Figure 9a and 10a). For example, for face identity representation, within-identity correlation would be the correlation between Damon face set 1 and Damon face set 2 or between DiCaprio face set 1 and DiCaprio face set 2, and between-identity correlation would be the correlation between Damon face set 1 and DiCaprio face set 1, between Damon face set 1 and DiCaprio face set 2, between Damon face set 2 and DiCaprio face set 1, or between Damon face set 2 and DiCaprio face set 2. If the average of all the within-identity correlations is higher than the average of all the between-identity correlations in an ROI, we would infer that abstract identity information was represented in that brain region. Correlation coefficients were Fisher-transformed to ensure normal distribution of the values before comparisons and statistical tests were conducted. All t-tests were two-tailed. When ANOVA was performed, Greenhouse-Geisser correction was applied if the sphericity assumption was violated.

In additional analyses, to ensure that voxel number differences did not contribute to the observed abstract object identity representation in superior IPS, we limited the total number of voxels in each ROI by selecting the top 50 most active voxels based on their average response amplitudes across all the stimulus conditions. In the large ROIs including superior IPS, LO, and PPA, we were able to select 50 voxels in each observer in each ROI. In the small ROIs including the right FFA and VWFA, we were able to select 50 voxels in the majority of the observers in each ROI (see Table 1 and Figure 13).

Behavioral and neural similarity measures of face identity

To construct behavioral similarity measures of face identity across the 8 actors, we obtained the search speed for each pairing of the 8 actors, totaling 28 pairs. The search speed for each pairing was averaged from trials with one actor serving as the target and the other one as the distractors and trials with the reverse assignment, as search speed from both types of trials reflected face identity similarity between the two actors. Furthermore, as search speed for target present and target absent trials were highly correlated in each observer (*Ps* < .026), these two types of trials were also combined (face similarity measure did not differ if only target present or target absent trials were included). A longer search speed would indicate a higher identity similarity between a given pair of actors, and a shorter search speed, on the other hand, would indicate a lower identity similarity between the two actors. The search speeds were extracted separately from each observer and then averaged across observers to form the group-level behavioral similarity measure of face identity.

To construct neural similarity measures of face identity across the 8 actors in each ROI, we Fisher transformed the correlation coefficients of neural response pattern correlation for each pair of actors, totaling 28 pairs. A higher correlation in this measure would indicate a higher identity similarity between a given pair of actors, and a lower correlation, on the other hand, would indicate a lower identity similarity between the two actors. These correlations were performed separately for each observer and then averaged across observers to form the group-level neural similarity measure of face identity separately for each ROI.

Behavioral and neural similarity measures of face identity across the 8 actors were then directly correlated for each ROI. If representations in a brain region reflected perception, then a high correlation between behavioral and neural similarity measures was expected (Kriegeskorte

et al., 2008). The significance of the correlation was evaluated using a permutation test (10,000 iterations) in which the values within the behavior and neural similarity measures were randomly shuffled and correlated for 10,000 iterations to derive the mean and SD of the baseline correlation value distribution.

4

Conclusion

In this dissertation, using univariate and multivariate fMRI analyses, I showed that superior IPS encodes various types of task-relevant visual information, from simple features such as shape to abstract object identities. In Chapter 1, I showed that the processing of task-irrelevant object shapes in superior IPS could be attenuated when their locations were known. The presence of distractor shapes increased fMRI response amplitudes in superior IPS during the VSTM encoding period when target encoding load was low. However, the presence of distractors did not increase response amplitudes when target locations were cued in advance, suggesting distractors were excluded from further processing. In Chapter 2, using MVPA, I found that shape information is represented in superior IPS, but only when such information is task-relevant. I tested shape decoding in superior IPS in four tasks that were perceptually same but had different task demands.

Object shape could be decoded in superior IPS, but only in the tasks that required such information, confirming that this region flexibly varies the type of information it encodes in a task-dependent manner. In Chapter 3, I showed that superior IPS represents not only basic visual featural information, but also abstract object identity. Specifically, in superior IPS, neural response patterns for face and car images of the same identity were more highly correlated than those for images of different identities, even when the perceptual features varied significantly within each identity. Importantly, I found the perceived similarity between identities measured by a behavioral visual search task significantly correlated with neural similarity measures in superior IPS, suggesting the representation in this region reflected the perceived identities.

In addition to the findings of flexible, task-dependent visual representation in superior IPS, these results also point out two important methodological issues for future research. First, the results of this dissertation suggest that the task used has a significant effect on the activation in superior IPS. The results in Chapter 1 and 2 showed that the same visual input was either encoded or not depending on task context. Therefore, to study the visual representation in IPS, one should use a task that directly requires the processing of that visual information. A passive task or a task that is orthogonal to independent variables of the study is not likely to recruit IPS regions.

Second, the present results suggest that the selection of task-relevant sub-region within IPS is critical. While I and others have found object representations, in particular VSTM information, within parietal cortex (Christophel et al., 2012; Christophel & Haynes, 2014; Todd & Marois, 2004; 2005; Xu, 2007; Xu & Chun, 2006; Xu & Jeong, in press), other

studies have not (Emrich, Riggall, Larocque, & Postle, 2013; Riggall & Postle, 2012). This could be due to functional heterogeneity in IPS. Examples of such heterogeneity include variations in shape sensitivity across topographic IPS regions (Konen & Kastner, 2008), and differing roles in object individuation and identification in superior and inferior IPS (Xu, 2009; Xu & Chun, 2006; 2009). Studies that have observed VSTM representations in parietal cortex either localized voxels that tracked VSTM capacity in individual observers or used a searchlight approach (Kriegeskorte, Goebel, & Bandettini, 2006) to locate informative voxels. Thus, these results suggest that it is important to isolate sub-regions within IPS that are directly involved in the task one wants to investigate.

Previously it was thought that featural and identity information was solely processed by the ventral visual pathway (Ungerleider & Haxby, 1994; Ungerleider & Mishkin, 1982). My results show a distinct role for the parietal cortex, in particular superior IPS, in this sort of processing. In the ventral pathway, different object categories, such as faces (Kanwisher et al., 1997), bodies (Downing, Jiang, Shuman, & Kanwisher, 2001), scenes (Epstein & Kanwisher, 1998), and so on, are processed in spatially separable clusters. However, to support flexible visual information processing, it would be optimal if all of these object categories can be processed in the same region depending on task demands. This dissertation shows that superior IPS is where a variety of visual information can be stored. My results show that superior IPS not only represents visual information from a variety of categories, but also at a variety of different levels, from simple shape (Chapter 1) to everyday objects (Chapter 2), and even abstract object identities (Chapter 3). These representations are dynamically modulated by task demands. All together, this

suggests that superior IPS may act as a general information storage region that, along with a network of other brain regions, supports flexible goal-directed behavior (Cole et al., 2013; Duncan, 2010; Fedorenko et al., 2013; Salazar, Dotson, Bressler, & Gray, 2012; Vincent et al., 2008).

Perhaps the most striking finding in this dissertation is that superior IPS can encode abstract object identities (Chapter 3). These results suggest that superior IPS does not just simply encode a "copy" of the visual input. Instead, this region seems to extract and integrate visual information to meet task demands, showing the unique role of human superior IPS in visual information processing.

The goal of this dissertation was to investigate the neural mechanisms that support flexible visual information processing. In conclusion, this dissertation provides evidence that superior IPS is a general visual information storage where various type of visual information is stored and integrated to support flexible goal-directed behavior.

Appendix to Chapter 3:

Abstract object identity representation in human superior intra-parietal sulcus

A.0 Perceptual differences among sets

Because we used photographs of face and car images as they appeared in the real world, minimal image processing was applied. Although we made sure that each face set contained a

similar range of variations in viewpoint, hair style, face expression and age and that each car set contained a similar range of variations in viewpoint, size and background scenes, lower-level perceptual differences such as luminance and image spatial frequency distribution could not be controlled for among the sets. Below we present three analyses showing that perceptual differences among sets could not account for the response patterns seen in superior IPS for abstract object identity representation in all three experiments.

Luminance difference among sets

In this analysis, we calculated the average luminance for each image in a set and compared whether sets differed in overall luminance. In Experiment 1, for the face images, face sets that shared identity were different from each other (*t*s > 2.26, *P*s < .036, independent samples t-test, two-tailed). Face sets that did not share identity were not significantly different from each other (*t*s < 1.62, *P*s > .12), except for one pair (Damon set 2 vs. DiCaprio set 2, $t_{(18)} = 3.51$, $P = .002$). Thus, luminance difference was greater within than between face identities. This made face images to be more similar when they did not share an identity than when they did, working against the finding of an identity effect in the superior IPS. For face names, no significant difference in luminance was found between all possible comparisons (*t*s < 1, *P*s > .72). In Experiment 2, for both car images and car names, there was no difference in luminance between sets that shared an identity and those that did not, *t*s < 1.67, *P*s > .11. In Experiment 3, no luminance difference was found between any of the face pairs (for all possible pairwise comparisons, *ts* < 1.25, *Ps* > .22).

Spatial frequency distribution differences among sets

In this analysis, we calculated spatial frequency distribution profile (i.e., the power at each spatial frequency) for each image in a set. We then used support vector machine, a linear classifier, to classify the images between sets based on this information to examine whether sets differed in spatial frequency distribution profiles. In Experiments 1 and 2, among all the comparisons made, the following yielded above chance level classification performance (ts > 2.44, Ps < .037, one-sample t-test, two-tailed): For face images, two out of the four between-identity comparisons; for face names, three out of the four between-identity comparisons; for car images, none; and for car names, three out of the four between-identity comparisons. These results showed that spatial frequency distribution envelope differed somewhat among the sets. Critically, although name sets were more similar when they shared an identity than when they did not, this similarity was not reflected in superior IPS response patterns for either face names or car names (see Figures 10 and 11). Additionally, although no difference was found among the car image sets, superior IPS response pattern still tracked car identity representation. Thus differences in spatial frequency distribution envelope did not seem to contribute to identity encoding in the superior IPS.

In Experiment 3, we performed the same spatial frequency distribution analysis with the face image sets. For within-identity comparisons, image set 1 of one actor was compared with image set 2 of the same actor, resulting in a total of 8 comparisons. For between-identity comparisons, image set 1 or 2 of one actor was compared with image set 1 or 2 of another actor, resulting in a total of 112 comparisons. Above chance classification performance was obtained in 1 out of the 8 within-identity comparisons ($t_{(9)} = 3$, $P = .015$) and in 40 out of the 224 between-

identity comparisons (*ts* > .244, *Ps* < .037). Although spatial frequency distribution differences between sets seemed to be greater between sets with different identities than those sharing the same identity, this difference was not registered by sensory regions as both LO and the right FFA showed similar correlations for within- and between-identity set comparisons. Given that sensory regions showed sensitivity to other perceptual differences among the image sets (see the analysis below on *Comparing same with different sets sharing an identity*), the insensitivity of sensory regions to spatial frequency distribution differences between sets suggests that they are unlikely to have contributed to the encoding of face identities in the brain.

Taken together, although there were some spatial frequency distribution differences among the images in different sets, these differences by themselves could not consistently account for the abstract object identity representation found in superior IPS across the three experiments.

Comparing same with different sets sharing an identity

One way to measure whether or not a brain region is sensitive to perceptual differences among sets is to compare its response patterns to odd and even runs of the same set of images with that from two different sets sharing the same identity. In other words, when set identity was held constant, because unique images were used in each set, if perceptual differences among the images were encoded by a brain region, its response pattern should be more similar to the same than to different set of images sharing an identity. Across the three experiments, as shown in Figure 14, the following ROIs showed a significantly higher correlation between identical sets than between sets sharing an identity: For face images in Experiment 1, none; for face names in

Experiment 1, the right FFA ($t_{(12)} = 2.45$, $P = .03$); for car images in Experiment 2, both PPA and VWFA (ts > 2.25, Ps < .044); and for car names in Experiment 2, none; for face images in Experiment 3, both LO and the right FFA (ts > 2.61, Ps < .026).

Thus, depending on the stimulus used, different ventral object processing regions showed different amount of sensitivity to perceptual or image differences between sets. Importantly, when identity was held constant, the superior IPS never differentiated between sets of images that were identical and those that were different. This provides further support that perceptual differences among sets did not modulate response pattern in superior IPS.

A.1 Identity encoding in LPFC, PPC, and ventral/temporal visual regions

Besides our functionally defined ROIs, the face oddball detection task also activated regions in LPFC, PPC and ventral/temporal visual cortices. To assess whether or not abstract face identity representation exists in these regions, face identity decoding was examine in these regions in Experiment 3.

To define LPFC, PPC and the ventral/temporal visual region ROIs, the continuous set of voxels showing a higher response to the face stimuli than to fixation in the main task (FDR q < .05, no more than 15 continuous voxels in each dimension was selected from the center of the activation) in the prefrontal cortex, posterior parietal cortex, and ventral region, respectively, were selected. To further refine our ROIs, voxels from frontal eye field, insular, and anterior cingulate cortex were excluded from LPFC, those from superior IPS were excluded from PPC,

and those from LO, the right FFA and early visual areas (localized by the contrast of scrambled objects minus intact objects in the LO localizer task) were excluded from the ventral/temporal visual region.

None of these brain regions showed identity representation: in LPFC, $t_{(10)} = 1.27$, $P = .23$; in PPC excluding superior IPS, $t_{(10)} = 1.61$, $P = .137$; and in ventral/temporal visual region excluding LO, the right FFA, and the early visual areas, $t_{(10)} = 1.01$, $P = .336$. Although identity decoding showed a trend in PPC with a p-value of .137, the PPC region we defined here contained voxels near superior IPS which might have contributed to some amount of identity representation. To examine this possibility, we relaxed the threshold and defined a superior IPS by adding roughly twice more number of adjacent voxels to the original superior IPS (on average 256 voxels) and then excluding this larger superior IPS from PPC. In this refined PPC, face identity decoding became less significant ($t_{(10)} = 1.38$, $P = .197$). This suggested that the voxels near superior IPS likely contributed to the trending face decoding in PPC when a smaller superior IPS was excluded.

Overall, these analyses showed that LPFC and ventral/temporal visual region did not contain abstract face identity representation. Moreover, not all visually responsive voxels in parietal cortex carry face identity representation, confirming the unique role of superior IPS in object identity representation.

a

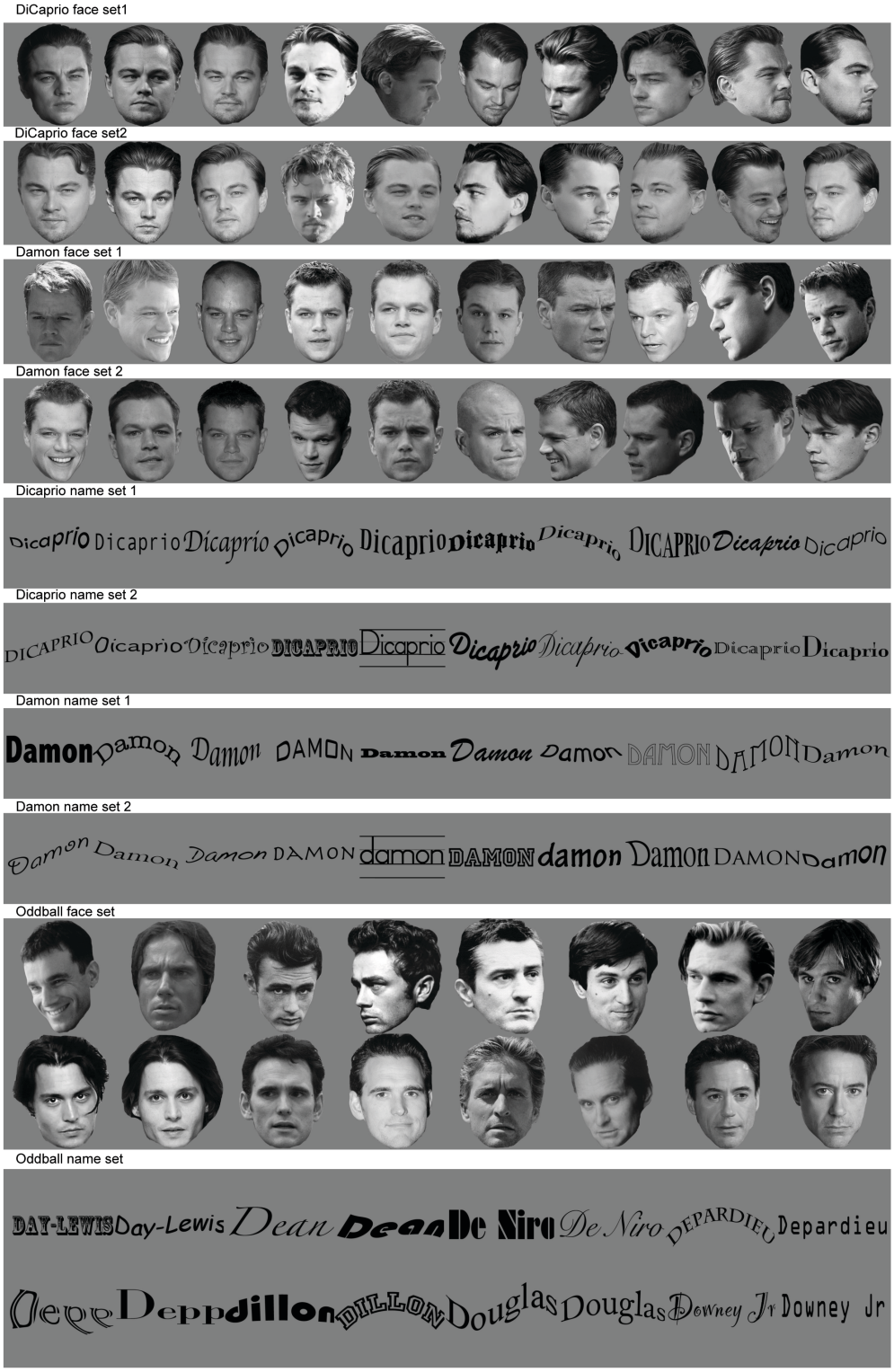

Figure 12. The full stimulus sets used, including both targets and oddballs, in **a.** Experiment 1, **b.** Experiment 2, and **c.** Experiment 3.

b

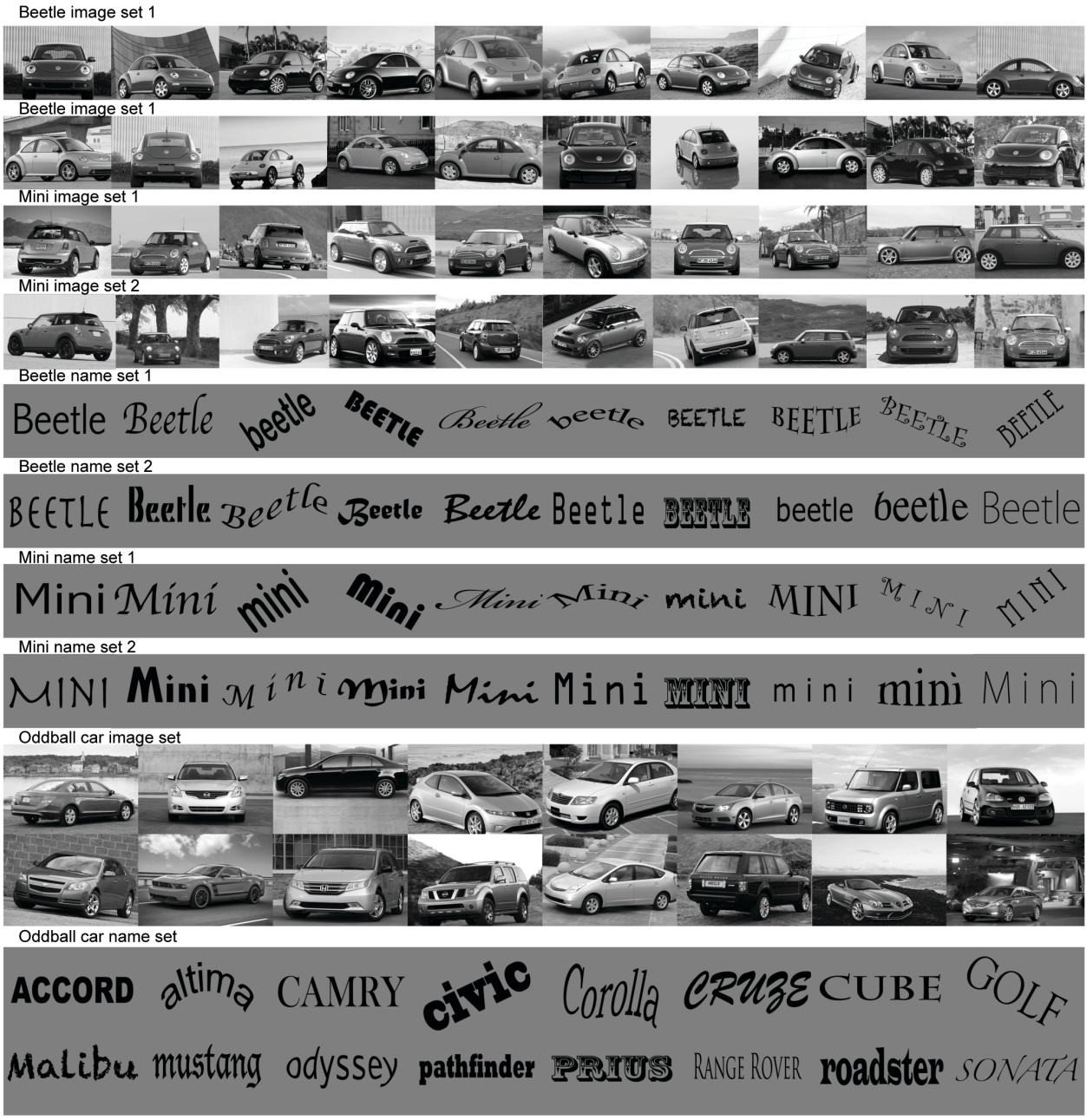

Figure 12. (Continued)

c

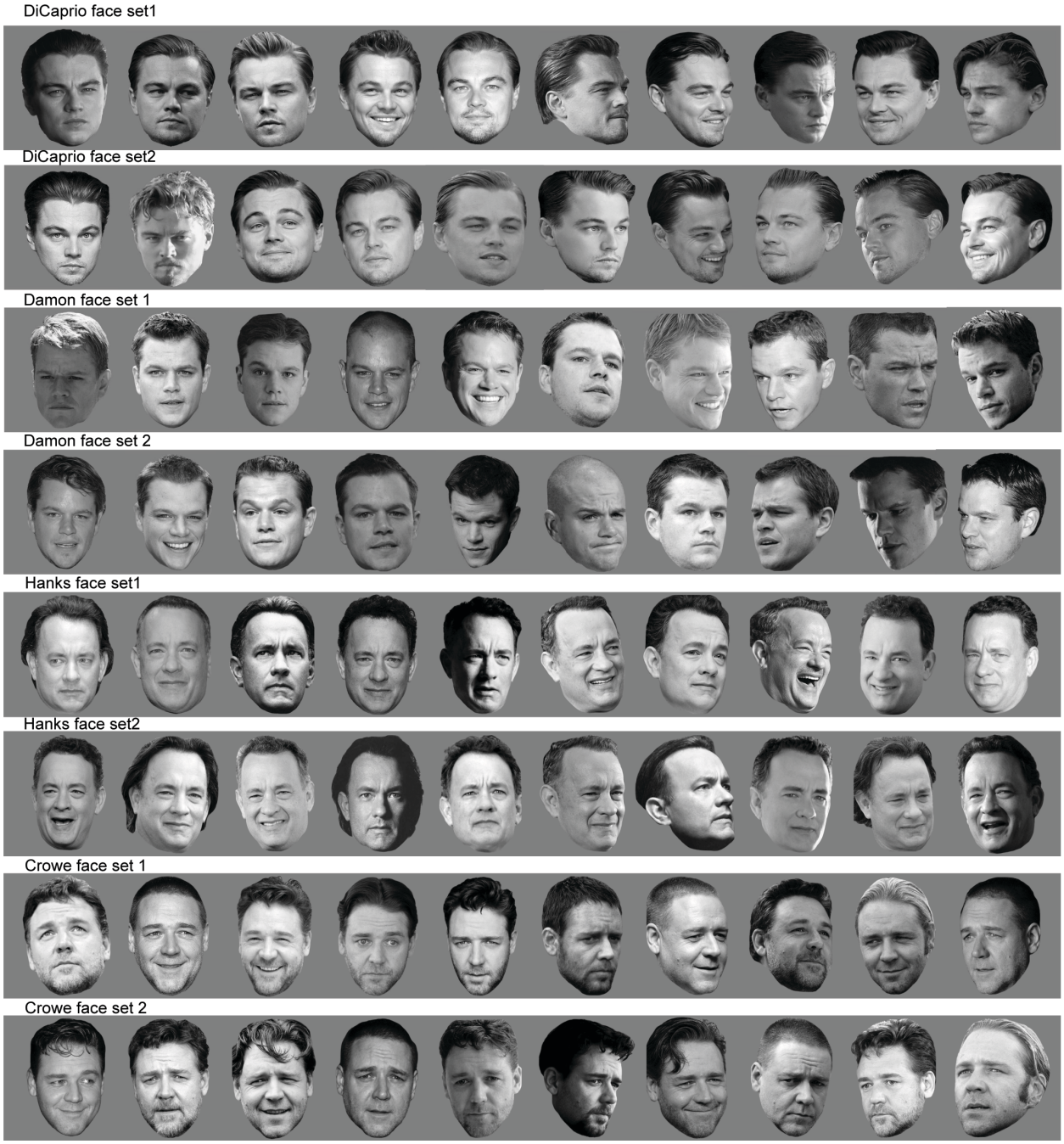

Figure 12. (Continued)

c

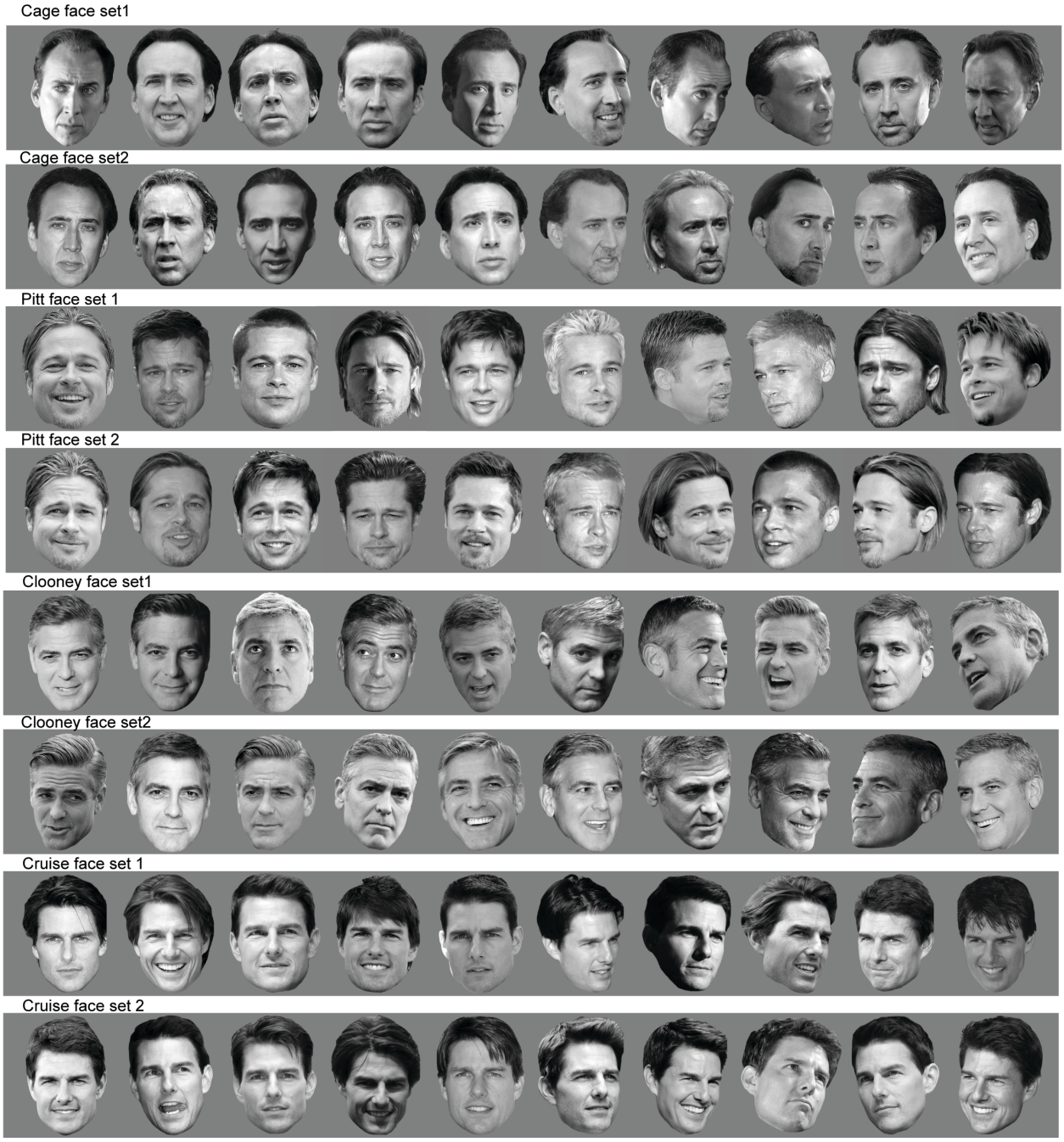

Figure 12. (Continued)

c

Figure 12. (Continued)

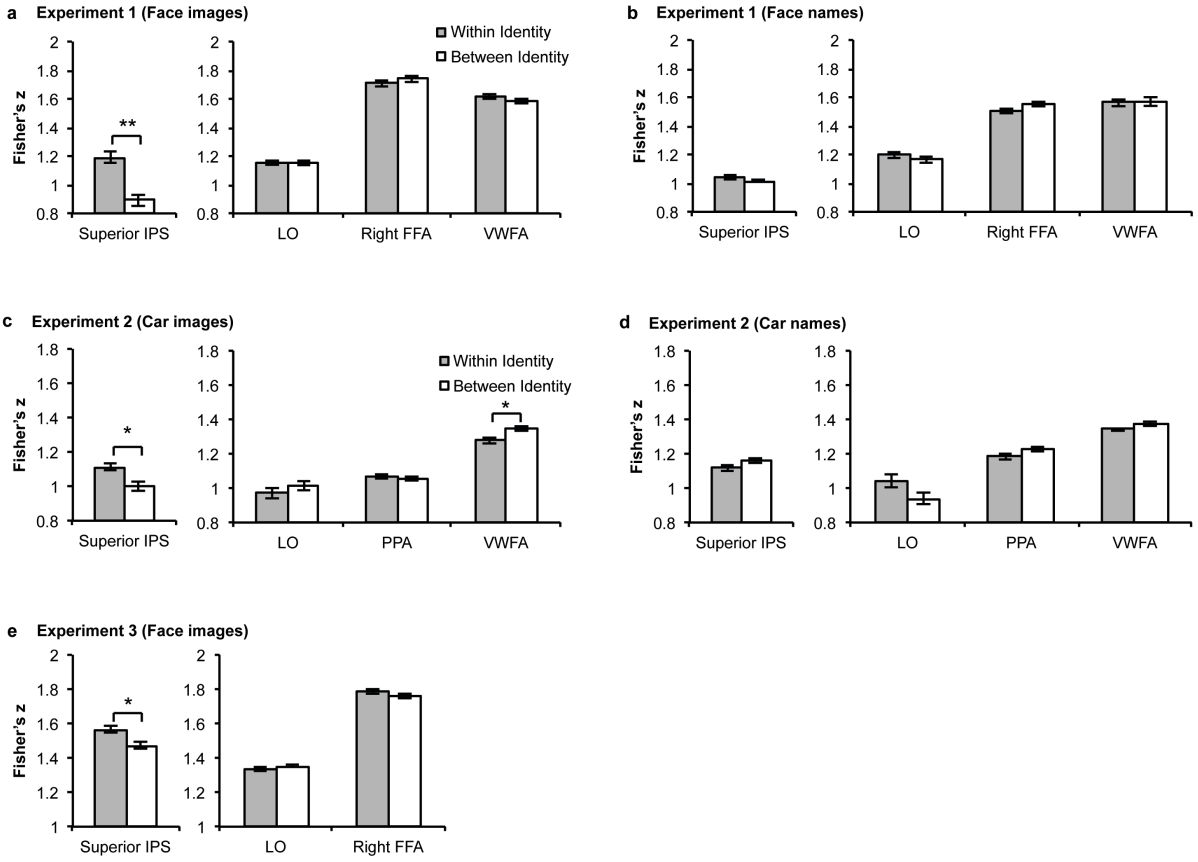

Figure 13. Identity representation in each ROI when no more than 50 voxels were included. Limiting the total number of voxels in each ROI did not affect the results and superior IPS was still the only region showing abstract object identity representations for faces and cars (* $P < .05$; ** $P < .01$. Error bars indicate within-subject standard error of the mean).

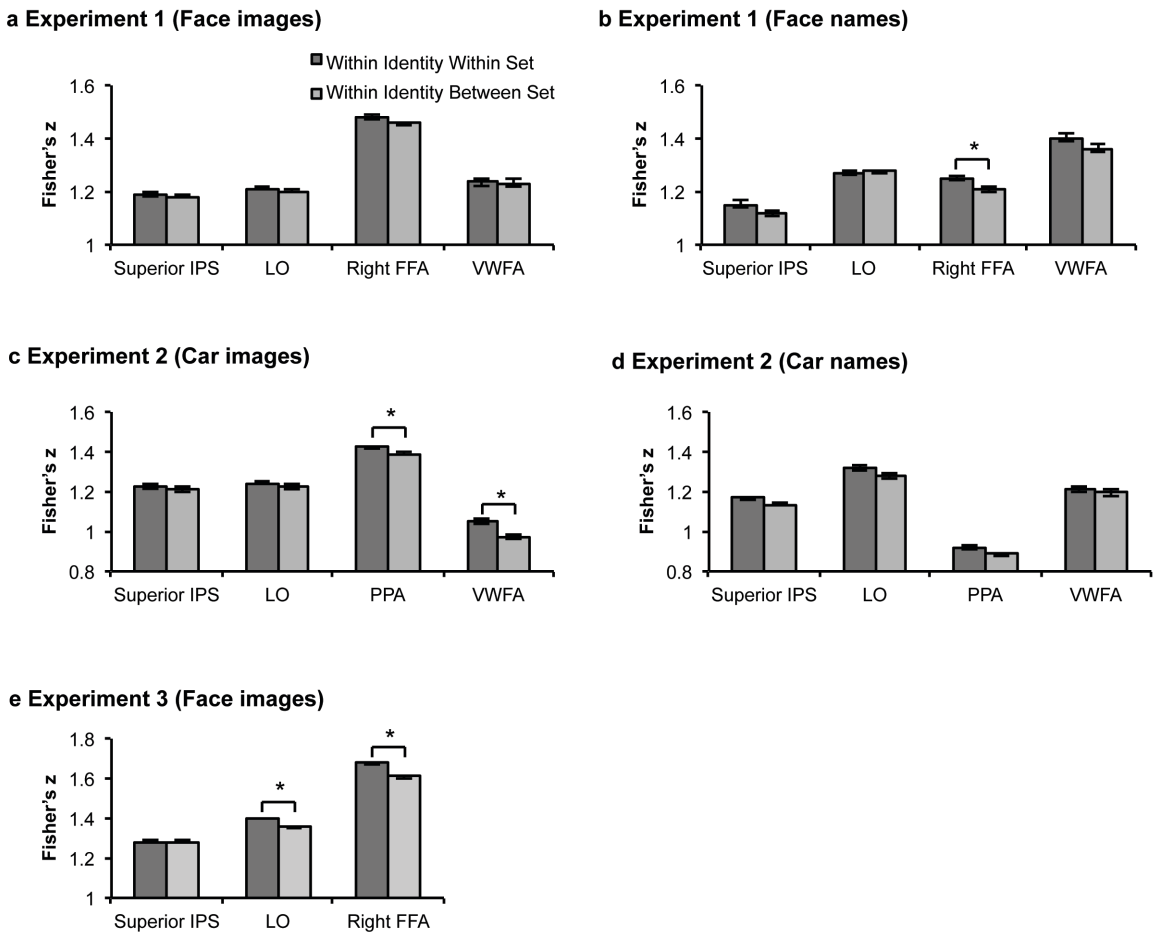

Figure 14. Neural response patterns to perceptual differences among sets sharing an identity in each ROI. While holding identity constant, correlation between odd and even runs of the same set was compared with correlation between two different sets sharing an identity. A higher within- than between-set correlation would indicate the encoding of perceptual or image differences between sets in a brain region. Depending on the stimulus used, different ventral object processing regions showed sensitivity to perceptual or image differences between sets. Importantly, with identity held constant, superior IPS never differentiated sets of images whether they were identical or different. This provides further support that perceptual differences among sets did not modulate response pattern in superior IPS.

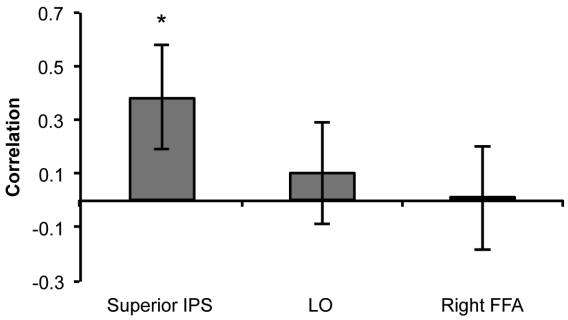

Figure 15. Correlation between behavioral and neural similarity measures of face identity when only up to 50 most responsive voxels were included in each ROI and when search speeds greater than 3 SD of the mean were removed. Only superior IPS, but not LO or the right FFA, showed a significant correlation.

References

Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science*, *15*, 106–111.

Anzellotti, S., Fairhall, S. L., & Caramazza, A. (2013). Decoding Representations of Face Identity That are Tolerant to Rotation. *Cerebral Cortex*, *24*, 1988-1995.

Baddeley, A. D. (1986). *Working memory*. Oxford University Press, USA.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433–436.

Broadbent, D. E. (1958). *Perception and communication.* New York, Pergamon Press.

Chang, C.-C., & Lin, C.-J. (2011). LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, *2*, 27:1–27:27.

Christophel, T. B., & Haynes, J.-D. (2014). Decoding complex flow-field patterns in visual working memory. *NeuroImage*, *91*, 43–51.

Christophel, T. B., Hebart, M. N., & Haynes, J.-D. (2012). Decoding the contents of visual short-term memory from human visual and parietal cortex. *Journal of Neuroscience*, *32*, 12983–12989.

Cohen, L., Dehaene, S., Naccache, L., Lehéricy, S., Dehaene-Lambertz, G., Hénaff, M. A., & Michel, F. (2000). The visual word form area: spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain*, *123 (Pt 2)*, 291–307.

Cohen, L., Lehéricy, S., Chochon, F., Lemer, C., Rivaud, S., & Dehaene, S. (2002). Language-specific tuning of visual cortex? Functional properties of the Visual Word Form Area. *Brain*, *125*, 1054–1069.

Colby, C. L., & Goldberg, M. E. (1999). Space and attention in parietal cortex. *Annual Review of Neuroscience*, *22*, 319–349.

Cole, M. W., Reynolds, J. R., Power, J. D., Repovs, G., Anticevic, A., & Braver, T. S. (2013). Multi-task connectivity reveals flexible hubs for adaptive task control. *Nature Neuroscience*, *16*, 1348–1355.

Corbetta, M., Miezin, F. M., Dobmeyer, S., Shulman, G. L., & Petersen, S. E. (1990). Attentional modulation of neural processing of shape, color, and velocity in humans. *Science*, *248*, 1556–1559.

Corkin, S. (1968). Acquisition of motor skill after bilateral medial temporal-lobe excision. *Neuropsychologia*, *6*, 255–265.

Corkin, S. (2002). What's new with the amnesic patient H.M.? *Nature Reviews. Neuroscience*, *3*, 153–160.

Corkin, S., Amaral, D. G., González, R. G., Johnson, K. A., & Hyman, B. T. (1997). H. M.'s medial temporal lobe lesion: findings from magnetic resonance imaging. *Journal of Neuroscience*, *17*, 3964–3979.

Cowan, N. (2001). The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *The Behavioral and Brain Sciences*, *24*, 87–114– discussion 114–85.

Cox, D. D., & Savoy, R. L. (2003). Functional magnetic resonance imaging (fMRI) "brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex. *NeuroImage*, *19*, 261–270.

De Renzi, E., Perani, D., Carlesimo, G. A., Silveri, M. C., & Fazio, F. (1994). Prosopagnosia can be associated with damage confined to the right hemisphere--an MRI and PET study and a review of the literature. *Neuropsychologia*, *32*, 893–902.

Deutsch, J. A., & Deutsch, D. (1963). Attention: Some Theoretical Considerations. *Psychological Review*, *70*, 80–90.

Dosenbach, N. U. F., Fair, D. A., Cohen, A. L., Schlaggar, B. L., & Petersen, S. E. (2008). A dual-networks architecture of top-down control. *Trends in Cognitive Sciences*, *12*, 99–105.

Downing, P. E., Jiang, Y., Shuman, M., & Kanwisher, N. G. (2001). A cortical area selective for visual processing of the human body. *Science*, *293*, 2470–2473.

Duncan, J. (2001). An adaptive coding model of neural function in prefrontal cortex. *Nature Reviews. Neuroscience*, *2*, 820–829.

Duncan, J. (2010). The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. *Trends in Cognitive Sciences*, *14*, 172–179.

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, *96*, 433–458.

Duncan, J., & Owen, A. M. (2000). Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends in Neurosciences*, *23*, 475–483.

Emrich, S. M., Riggall, A. C., Larocque, J. J., & Postle, B. R. (2013). Distributed patterns of activity in sensory cortex reflect the precision of multiple items maintained in visual short-term memory. *Journal of Neuroscience*, *33*, 6516–6523.

Epstein, R. A., & Kanwisher, N. G. (1998). A cortical representation of the local visual environment. *Nature*, *392*, 598–601.

Fedorenko, E., Duncan, J., & Kanwisher, N. G. (2013). Broad domain generality in focal regions of frontal and parietal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *110*, 16616–16621.

Fitzgerald, J. K., Freedman, D. J., & Assad, J. A. (2011). Generalized associative representations in parietal cortex. *Nature Neuroscience*, *14*, 1075–1079.

Fitzgerald, J. K., Freedman, D. J., Fanini, A., Bennur, S., Gold, J. I., & Assad, J. A. (2013). Biased associative representations in parietal cortex. *Neuron*, *77*, 180–191.

Fitzgerald, J. K., Swaminathan, S. K., & Freedman, D. J. (2012). Visual categorization and the parietal cortex. *Frontiers in Integrative Neuroscience*, *6*, 18.

Freedman, D. J., & Assad, J. A. (2006). Experience-dependent representation of visual categories in parietal cortex. *Nature*, *443*, 85–88.

Freedman, D. J., & Assad, J. A. (2009). Distinct encoding of spatial and nonspatial visual information in parietal cortex. *Journal of Neuroscience*, *29*, 5671–5680.

Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2001). Categorical representation of visual stimuli in the primate prefrontal cortex. *Science*, *291*, 312–316.

Goldman-Rakic, P. S. (1995). Cellular basis of working memory. *Neuron*, *14*, 477–485.

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, *15*, 20–25.

Gottlieb, J. P., & Balan, P. (2010). Attention as a decision in information space. *Trends in Cognitive Sciences*, *14*, 240–248.

Gottlieb, J. P., Kusunoki, M., & Goldberg, M. E. (1998). The representation of visual salience in monkey parietal cortex. *Nature*, *391*, 481–484.

Grill-Spector, K., Kushnir, T., Edelman, S., Itzchak, Y., & Malach, R. (1998). Cue-invariant activation in object-related areas of the human occipital lobe. *Neuron*, *21*, 191–202.

Grill-Spector, K., Kushnir, T., Hendler, T., & Malach, R. (2000). The dynamics of object-selective activation correlate with recognition performance in humans. *Nature Neuroscience*, *3*, 837–843.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, *293*, 2425–2430.

Jeong, S. K., & Xu, Y. (2013). Neural Representation of Targets and Distractors during Object Individuation and Identification. *Journal of Cognitive Neuroscience*, *25*, 117–126.

Kahneman, D., Treisman, A. M., & Gibbs, B. J. (1992). The reviewing of object files: object-specific integration of information. *Cognitive Psychology*, *24*, 175–219.

Kamitani, Y., & Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nature Neuroscience*, *8*, 679–685.

Kanwisher, N. G., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, *17*, 4302–4311.

Konen, C. S., & Kastner, S. (2008). Two hierarchically organized neural systems for object information in human visual cortex. *Nature Neuroscience*, *11*, 224–231.

Kourtzi, Z., & Kanwisher, N. G. (2000). Cortical regions involved in perceiving object shape. *Journal of Neuroscience*, *20*, 3310–3318.

Kourtzi, Z., & Kanwisher, N. G. (2001). Representation of perceived object shape by the human lateral occipital complex. *Science*, *293*, 1506–1509.

Kravitz, D. J., Kriegeskorte, N., & Baker, C. I. (2010). High-level visual object representations are constrained by position. *Cerebral Cortex*, *20*, 2916–2925.

Kravitz, D. J., Saleem, K. S., Baker, C. I., & Mishkin, M. (2011). A new neural framework for visuospatial processing. *Nature Reviews. Neuroscience*, *12*, 217–230.

Kravitz, D. J., Saleem, K. S., Baker, C. I., Ungerleider, L. G., & Mishkin, M. (2013). The ventral visual pathway: an expanded neural framework for the processing of object quality. *Trends in Cognitive Sciences*, *17*, 26–49.

Kriegeskorte, N., Formisano, E., Sorger, B., & Goebel, R. (2007). Individual faces elicit distinct response patterns in human anterior temporal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *104*, 20600–20605.

Kriegeskorte, N., Goebel, R., & Bandettini, P. A. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 3863–3868.

Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., et al. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron*, *60*, 1126–1141.

Lavie, N. (2005). Distracted and confused?: selective attention under load. *Trends in Cognitive Sciences*, *9*, 75–82.

Lavie, N., & Tsal, Y. (1994). Perceptual load as a major determinant of the locus of selection in visual attention. *Perception & Psychophysics*, *56*, 183–197.

Lavie, N., Hirst, A., de Fockert, J. W., & Viding, E. (2004). Load theory of selective attention and cognitive control. *Journal of Experimental Psychology: General*, *133*, 339–354.

Liu, T., Hospadaruk, L., Zhu, D. C., & Gardner, J. L. (2011). Feature-specific attentional priority signals in human cortex. *Journal of Neuroscience*, *31*, 4484–4495.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*, 279–281.

Luck, S. J., Vogel, E. K., & Shapiro, K. L. (1996). Word meanings can be accessed but not reported during the attentional blink. *Nature*, *383*, 616–618.

Malach, R., Reppas, J. B., Benson, R. R., Kwong, K. K., Jiang, H., Kennedy, W. A., et al. (1995). Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *92*, 8135–8139.

Marois, R., & Ivanoff, J. (2005). Capacity limits of information processing in the brain. *Trends in Cognitive Sciences*, *9*, 296–305.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual*

Review of Neuroscience, *24*, 167–202.

Milner, A. D., Perrett, D. I., Johnston, R. S., Benson, P. J., Jordan, T. R., Heeley, D. W., et al. (1991). Perception and action in 'visual form agnosia'. *Brain*, *114 (Pt 1B)*, 405–428.

Milner, B., Corkin, S., & Teuber, H. L. (1968). Further analysis of the hippocampal amnesic syndrome: 14-year follow-up study of H.M. *Neuropsychologia*, *6*, 215–234.

Moray, N. (1959). Attention in dichotic listening: Affective cues and the influence of instructions. *Quarterly journal of experimental psychology*, *11*, 56–60.

Nestor, A., Plaut, D. C., & Behrmann, M. (2011). Unraveling the distributed neural code of facial identity through spatiotemporal pattern analysis. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, 9998–10003.

Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences*, *10*, 424–430.

O'Craven, K. M., Downing, P. E., & Kanwisher, N. G. (1999). fMRI evidence for objects as the units of attentional selection. *Nature*, *401*, 584–587.

Phillips, W. A. (1974). On the distinction between sensory storage and short-term visual memory. *Perception & Psychophysics*, *16*, 283–290.

Pinsk, M. A., Doniger, G. M., & Kastner, S. (2004). Push-pull mechanism of selective attention in human extrastriate cortex. *Journal of Neurophysiology*, *92*, 622–629.

Posner, M. I. (1980). Orienting of attention. *Quarterly journal of experimental psychology*, *32*, 3–25.

Pylyshyn, Z. (1989). The role of location indexes in spatial perception: a sketch of the FINST spatial-index model. *Cognition*, *32*, 65–97.

Pylyshyn, Z. (1994). Some primitive mechanisms of spatial attention. *Cognition*, *50*, 363–384.

Riggall, A. C., & Postle, B. R. (2012). The relationship between working memory storage and elevated activity as measured with functional magnetic resonance imaging. *Journal of Neuroscience*, *32*, 12990–12998.

Salazar, R. F., Dotson, N. M., Bressler, S. L., & Gray, C. M. (2012). Content-Specific Fronto-Parietal Synchronization During Visual Working Memory. *Science*, *338*, 1097–1100.

Scholl, B. J. (2001). Objects and attention: the state of the art. *Cognition*, *80*, 1–46.

Schwarzlose, R. F., Swisher, J. D., Dang, S., & Kanwisher, N. G. (2008). The distribution of category and location information across object-selective regions in human visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 4447–4452.

Scoville, W. B., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of Neurology, Neurosurgery, and Psychiatry*, *20*, 11–21.

Serences, J. T., Ester, E. F., Vogel, E. K., & Awh, E. (2009). Stimulus-specific delay activity in human primary visual cortex. *Psychological Science*, *20*, 207–214.

Sereno, A. B., & Maunsell, J. H. (1998). Shape selectivity in primate lateral intraparietal cortex. *Nature*, *395*, 500–503.

Stoet, G., & Snyder, L. H. (2004). Single neurons in posterior parietal cortex of monkeys encode cognitive set. *Neuron*, *42*, 1003–1012.

Swaminathan, S. K., & Freedman, D. J. (2012). Preferential encoding of visual categories in parietal cortex compared with prefrontal cortex. *Nature Neuroscience*, *15*, 315–320.

Talairach, J., & Tournoux, P. (1988). *Co-Planar Stereotaxic Atlas of the Human Brain*. Thieme.

Thompson, R., & Duncan, J. (2009). Attentional modulation of stimulus representation in human fronto-parietal cortex. *NeuroImage*, *48*, 436–448.

Todd, J. J., & Marois, R. (2004). Capacity limit of visual short-term memory in human posterior parietal cortex. *Nature*, *428*, 751–754.

Todd, J. J., & Marois, R. (2005). Posterior parietal cortex activity predicts individual differences in visual short-term memory capacity. *Cognitive, Affective & Nehavioral Neuroscience*, *5*, 144–155.

Torralbo, A., & Beck, D. M. (2008). Perceptual-load-induced selection as a result of local competitive interactions in visual cortex. *Psychological Science*, *19*, 1045–1050.

Toth, L. J., & Assad, J. A. (2002). Dynamic coding of behaviourally relevant stimuli in parietal cortex. *Nature*, *415*, 165–168.

Tsao, D. Y., Moeller, S., & Freiwald, W. A. (2008). Comparing face patch systems in macaques and humans. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 19514–19519.

Ungerleider, L. G., & Haxby, J. V. (1994). 'What' and "where" in the human brain. *Current Opinion in Neurobiology*, *4*, 157–165.

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale, & R. J. W. Mansfield (Eds.), *Analysis of Visual Behavior* (pp. 549–586). Cambridge: MIT press.

Ungerleider, L. G., Courtney, S. M., & Haxby, J. V. (1998). A neural system for human visual working memory. *Proceedings of the National Academy of Sciences of the United States of America*, *95*, 883–890.

Van Orden, G. C. (1991). Phonologic mediation is fundamental to reading. In B. Derek & G. W. Humphreys (Eds.), *Basic processes in reading* (pp. 77–103). Hillsdale, NJ: Erlbaum.

Vincent, J. L., Kahn, I., Snyder, A. Z., Raichle, M. E., & Buckner, R. L. (2008). Evidence for a Frontoparietal Control System Revealed by Intrinsic Functional Connectivity. *Journal of Neurophysiology*, *100*, 3328–3342.

Vogel, E. K., & Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature*, *428*, 748–751.

Vogel, E. K., McCollough, A. W., & Machizawa, M. G. (2005). Neural measures reveal individual differences in controlling access to working memory. *Nature*, *438*, 500–503.

Woolgar, A., Hampshire, A., Thompson, R., & Duncan, J. (2011). Adaptive coding of task-relevant information in human frontoparietal cortex. *Journal of Neuroscience*, *31*, 14592–14599.

Xu, Y. (2007). The role of the superior intraparietal sulcus in supporting visual short-term memory for multifeature objects. *Journal of Neuroscience*, *27*, 11676–11686.

Xu, Y. (2008). Representing connected and disconnected shapes in human inferior intraparietal sulcus. *NeuroImage*, *40*, 1849–1856.

Xu, Y. (2009). Distinctive neural mechanisms supporting visual object individuation and identification. *Journal of Cognitive Neuroscience*, *21*, 511–518.

Xu, Y. (2010). The neural fate of task-irrelevant features in object-based processing. *Journal of Neuroscience*, *30*, 14020–14028.

Xu, Y., & Chun, M. M. (2006). Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature*, *440*, 91–95.

Xu, Y., & Chun, M. M. (2007). Visual grouping in human parietal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *104*, 18766–18771.

Xu, Y., & Chun, M. M. (2009). Selecting and perceiving multiple visual objects. *Trends in Cognitive Sciences*, *13*, 167–174.

Xu, Y., & Jeong, S. K. (in press). The contribution of human superior intra-parietal sulcus to visual short-term memory and perception. In P. Jolicoeur & J. Martinez-Trujillo (Eds.), *Mechanisms of Sensory Working Memory: Attention and Performance XXV*.

Yi, D.-J., Woodman, G. F., Widders, D., Marois, R., & Chun, M. M. (2004). Neural fate of ignored stimuli: dissociable effects of perceptual and working memory load. *Nature Neuroscience*, *7*, 992–996.

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*, 233–235.